

Anderson, S., Barna, N., Bastien, V., Bloom, T., Boguslavskiy, L., Boukhgalter, B., Brown, A., Camarata, J., Campopiano, A., Chang, J., Chazaro, B., Choepel, Y., Collangelo, M., Collins, S., Collimore, A., Cook, A., Cooke, P., DeArellano, K., Dewar, K., Diaz, J.S., Dodge, S., Faro, S., Ferreira, P., Fitzgerald, M., FitzHugh, W., Gage, D., Galagan, J., Gardyna, S., Ginde, S., Gord, S., Goyette, M., Graham, L., Grand-Pierre, N., Hagos, B., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C., Kamat, A., Karatas, A., Kells, C., LaRoque, K., Lamazares, R., Landers, T., Lehoczyk, J., Levine, R., Lindblad-Toh, K., Liu, G., MacLean, C., Macdonald, P., Major, J., Marquis, N., Matthews, C., McCarthy, M., McEwan, P., McKernan, K., Meldrim, J., Meneus, L., Mihova, T., Mlenga, V., Murphy, T., Naylor, J., Nguyen, C., Nicol, R., Norbu, C., Norman, C.H., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Peterson, K., Phunkhang, P., Pierre, N., Pollara, V., Raymond, C., Retta, R., Rieback, M., Riley, R., Rise, C., Rogov, P., Roman, J., Rosetti, M., Roy, A., Santos, R., Schauer, S., Schupback, R., Seaman, S., Severy, P., Spencer, B., Strange-Thomann, N., Stojanovic, N., Strauss, N., Subramanian, A., Talamas, J., Tesfaye, S., Theodore, J., Topham, K., Travers, M., Travis, N., Triggillo, J., Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W. J., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

Direct Submission
Submitted (09-JUN-2002) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
3 (bases 1 to 194420)

TITLE JOURNAL REFERENCE AUTHORS

Bitren, B., Nussbaum, C., Lander, E., Abouelleil, A., Allen, N., Anderson, M., Arachchi, H.M., Barna, N., Bastien, V., Bloom, T., Boguslavskiy, L., Boukhgalter, B., Camarata, J., Chang, J., Choepel, Y., Collimore, A., Cook, A., Cooke, P., Corum, B., DeArellano, K., Diaz, J.S., Dodge, S., Dooley, K., Dorris, L., Erickson, J., Faro, S., Ferreira, P., Fitzgerald, M., Gage, D., Galagan, J., Gardyna, S., Graham, L., Grand-Pierre, N., Hafez, N., Hagopian, D., Hagos, B., Hall, J., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C., Kamat, A., Karatas, A., Kells, C., Landers, T., Levine, R., Lindblad-Toh, K., Liu, X., Lui, A., Mabbitt, R., MacLean, C., Macdonald, P., Major, J., Manning, J., Matthews, C., McCarthy, M., Meldrim, J., Meneus, L., Mihova, T., Mlenga, V., Murphy, T., Naylor, J., Nguyen, C., Nicol, R., Norbu, C., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Peterson, K., Phunkhang, P., Pierre, N., Rachupka, A., Ramasamy, U., Raymond, C., Retta, R., Rise, C., Rogov, P., Roman, J., Schauer, S., Schupback, R., Seaman, S., Severy, P., Smith, C., Spencer, B., Strange-Thomann, N., Stojanovic, N., Stubbs, M., Talamas, J., Tesfaye, S., Theodore, J., Topham, K., Travers, M., Vassiliev, H., Venkataraman, V.S., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

Direct Submission
Submitted (18-DEC-2003) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
On Dec 18, 2003 this sequence version replaced g1:38259237.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

TITLE JOURNAL COMMENT

----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIER
Web site: <http://www-seq.wi.mit.edu>
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L23552
Center clone name: 218_L_23

* NOTE: This is a 'working draft' sequence. It currently consists of 2 contigs. Gaps between the contigs are represented as runs of N. The order of the pieces is believed to be correct as given, however the sizes of the gaps between them are based on estimates that have been provided by the submittor.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
* 1 115229: contig of 115229 bp in length
* 115230 115329: gap of 100 bp
* 115330 194420: contig of 79091 bp in length.

FEATURES source

Location/Qualifiers
1. 194420
/organism="Mus musculus"
/mol_type="genomic DNA"
/db_xref="taxon:10090"
/chromosome="15"
/map="15"
/clones="RP23-218L23"
/clone_lib="RPC1-23 Female Mouse BAC"

ORIGIN

Alignment Scores:
Pred. No.: 374 Length: 194420
Score: 58.00 Matches: 11
Percent Similarity: 70.59% Conservatave: 1
Best Local Similarity: 64.71% Mismatches: 5
Query Match: 58.59% Indels: 0
DB: 2 Gaps: 0

US-10-799-747-116 (1-20) x AC124133 (1-194420)

Oy 1 MetaLalaHisSerValLeuSerPheLeuLeuTrpThrProTyraAlaLeu 17
Db 29103 ATGAACCTTCACAAGTCCTTCTATCTCTCTGGACACCCATTAGTCTC 29053

RESULT 8

AC096099/c

LOCUS AC096099 244669 bp DNA linear HTG 13-NOV-2002
DEFINITION Rattus norvegicus clone CH230-27M23, WORKING DRAFT SEQUENCE.
ACCESSION AC096099
VERSION AC096099.6 GI:24941648
KEYWORDS HTG; HTGS PHASE2; HTGS DRAFT; HTGS_FULLTOP.
SOURCE Rattus norvegicus (Norway rat)
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.

REFERENCE

AUTHORS

Muzny, D., Marté, Metzker, M., Lee, Abramzon, S., Adams, C., Alder, J., Allen, C., Allen, H., Alsbrooks, S., Amin, A., Anguiano, D., Aryalabechi, V., Aoyagi, A., Ayodeji, M., Baca, E., Baden, H., Baldwin, D., Bandaranaike, D., Barber, M., Barnstead, M., Benahmed, F., Bialski, K., Blair, J., Blankenburg, K., Blyth, P., Brown, M., Bryant, N., Buhaq, C., Burch, P., Burrell, K., Calderon, E., Cardenas, V., Carter, K., Cavazos, I., Ceasar, H., Center, A., Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Y., Chen, Z., Chu, J., Cleveland, C., Cockrell, R., Cox, C., Coyle, M., Cree, A., D'Souza, L., Davila, M.L., Davis, C., Davy-Carroll, L., De Anda, C., Dederich, D., Delgado, O., Denson, S., Deramo, C., Ding, Y., Dinh, H., Divya, K., Draper, H., Dugan-Rocha, S., Dunn, A., Durbin, K., Duval, B., Eaves, K., Egan, A., Escotto, M., Eugene, C., Evans, C.A., Falls, T., Fan, G., Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P., Fraser, C.M., Gabisi, A., Ganta, R., Garcia, A., Garner, T., Garza, M., Gebregorgis, E., Geer, K., Gill, R., Grady, M., Guerra, W., Guevara, W., Gunaratne, P., Haaland, W., Hamill, C., Hamilton, C., Hamilton, K., Harvey, Y., Havlak, P., Hawes, A., Henderson, N., Hernandez, J., Hernandez, R., Hines, S., Hladun, S.L., Hodgson, A., Hogues, M., Hollins, B., Howells, S., Hulyk, S., Hume, J., Idlebird, D., Jackson, A., Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolivet, A., Karpachy, S., Kelly, S., Kelly, S., Khan, Z., King, L., Kovar, C., Kowis, C., Kraft, C.L., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, J., Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J., Lorensu, L., Louised, H., Lozano, R., Lu, X., Ma, J., Maheshwari, M., Mahindaratne, M., Mahmoud, M., Malloy, K., Mangum, A., Mangum, B., Mapua, P., Martin, K., Martin, R., Martinez, E., Mawhney, S., McLeod, M.P., McNeill, T.Z., Meenen, E., Mlloavljjevic, A., Miner, G., Minja, E., Montemayor, J., Moore, S., Morgan, M., Morris, K., Morris, S., Munidasa, M., Murphy, M., Nair, L., Nankervis, C., Neal, D., Newton, N., Nguyen, N., Norris, S., Nwakaemeh, O., Okwuonu, G., Olarnpunsagoon, A., Pal, S., Parks, K., Pasternak, S., Paul, H., Perez, A., Perez, L., Pfannkoch, C., Plopper, F., Poindexter, A., Popovic, D., Primus, E., Pu, L.-L., Puazo, M., Quiroz, J., Rachlin, E., Reeves, K., Regier, M.A., Reigh, R.,

Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, F.,
 Rivers, C., Rodkey, T., Rojas, A., Rose, M., Rose, R., Ruiz, S.J.,
 Sanders, W., Savary, G., Scherer, S., Scott, G., Shatsman, S., Shen, H.,
 Shetty, J., Shvartsbeyn, A., Sisson, I., Sitter, C.D., Smajls, D.,
 Sneed, A., Sodergren, E., Song, X.-Z., Sorelle, R., Sosa, J.,
 Steimle, M., Strong, R., Sutton, A., Svatek, A., Tabor, P., Taylor, C.,
 Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Umani, K.,
 Valas, R., Vera, V., Villasana, D., Waldron, L., Walker, B., Wang, J.,
 Wang, Q., Wang, S., Warren, J., Warren, R., Wei, X., White, F.,
 Williams, G., Willson, R., Wleciyk, R., Wooden, H., Worley, K.,
 Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V.,
 Yu, F., Zhang, J., Zhou, J., Zhou, X., Zhao, S., Dunn, D., von
 Niederhausern, A., Weiss, R., Smith, D.R., Holt, R.A., Smith, H.O.,
 Weinstein, G. and Gibbs, R.A.

TITLE
 JOURNAL
 REFERENCE
 AUTHORS
 TITLE
 JOURNAL

Submitted (17-SEP-2001) Human Genome Sequencing Center, Department
 of Molecular and Human Genetics, Baylor College of Medicine, One
 Baylor Plaza, Houston, TX 77030, USA

REFERENCE
 AUTHORS
 TITLE
 JOURNAL

3 (bases 1 to 244669)
 Rat Genome Sequencing Consortium.

COMMENT

Submitted (13-NOV-2002) Human Genome Sequencing Center, Department
 of Molecular and Human Genetics, Baylor College of Medicine, One
 Baylor Plaza, Houston, TX 77030, USA
 On Nov 13, 2002 this sequence version replaced gi:22772123.
 The sequence in this assembly is a combination of BAC based reads
 and whole genome shotgun sequencing reads assembled using Atlas
 (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described
 in the feature table below represents a scaffold in the Atlas
 assembly (a 'contig-scaffold'). Within each contig-scaffold,
 individual sequence contigs are ordered and oriented, and separated
 by sized gaps filled with Ns to the estimated size. The sequence
 may extend beyond the ends of the clone and there may be sequence
 contigs within a contig-scaffold that consist entirely of whole
 genome shotgun sequence reads. Both end sequences and whole genome
 shotgun sequence only contigs will be indicated in the feature
 table.

----- Genome Center
 Center: Baylor College of Medicine
 Center code: BCM
 Web site: http://www.hgsc.bcm.tmc.edu/
 Contact: hgsc-help@bcm.tmc.edu
 ----- Project Information
 Center project name: GEKW
 Center clone name: CH230-27M23

----- Summary Statistics
 Assembly program: Phrap; version 0.990329
 Consensus quality: 215199 bases at least Q40
 Consensus quality: 218145 bases at least Q30
 Consensus quality: 219884 bases at least Q20
 Estimated insert size: 219668; sum-of-contigs estimation
 Quality coverage: 7x in Q20 bases; sum-of-contigs estimation

* NOTE: Estimated insert size may differ from sequence length
 (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 1 contigs. Gaps between the contigs
 * are represented as runs of N. The order of the pieces
 * is believed to be correct as given, however the sizes
 * of the gaps between them are based on estimates that have
 * provided by the submitter.
 * This sequence will be replaced
 * by the finished sequence as soon as it is available and
 * the accession number will be preserved.
 * the accession number will be preserved.

FEATURES
 source

1 .244669
 Location/Qualifiers
 /organism="Rattus norvegicus"
 /mol_type="genomic DNA"

/db_xref="taxon:10116"
 /clone="CH230-27M23"
 1..1344
 /note="wgs contig"
 misc_feature
 4991..6114
 /note="wgs contig"
 misc_feature
 92124..93176
 /note="wgs contig"
 misc_feature
 243409..244669
 /note="wgs contig"

ORIGIN

Alignment Scores:
 Pred. No.: 485 Length: 244669
 Score: 58.00 Matches: 9
 Percent Similarity: 81.25% Conservative: 4
 Best Local Similarity: 56.25% Mismatches: 3
 Query Match: 58.59% Indels: 0
 DB: 2 Gaps: 0

US-10-799-747-116 (1-20) x AC096099 (1-244669)

Qy 1 MetaAlaHisSerValLeuSerPheLeuLeuThrProTyzAla 16

Db 112297 ATGTGTGCACACTCTGTCTTTCTTTTGTGATTGGTTACCTCACTCA 112250

RESULT 9

AC140313/c

LOCUS

DEFINITION

AC140313

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

Center: Washington University Genome Sequencing Center

Center code: WUGSC

Web site: http://genome.wustl.edu/gsc/index.shtml

Contact: submissions@wustl.edu

----- Project Information

Center project name: M_BB0324C06

----- Summary Statistics

Sequencing vector: M13; 0%

Chemistry: Dye-primer ET; 0% of reads

Assembly program: Phrap; version 0.990319

Consensus quality: 242982 bases at least Q40

Consensus quality: 243412 bases at least Q30

Consensus quality: 243721 bases at least Q20

* NOTE: This is a 'working draft' sequence. It currently

* consists of 7 contigs. The true order of the pieces

* is not known and their order in this sequence record is

* arbitrary. Gaps between the contigs are represented as

* runs of N, but the exact sizes of the gaps are unknown.

* This record will be updated with the finished sequence

* as soon as it is available and the accession number will

* be preserved.

1 1085: contig of 1085 bp in length

* 1086 1185: gap of unknown length

* 1186 1194: contig of 10759 bp in length

* 1194 1204: gap of unknown length

* 1204 1245 12044: gap of unknown length

* 1245 30235: contig of 18181 bp in length

* 30235 30326: gap of unknown length

* 30326 63384: contig of 33059 bp in length

* 63384 63485 63484: gap of unknown length

* 63485 117854: contig of 54370 bp in length

* 117854 166434: gap of unknown length

* 166434 166435: contig of 48480 bp in length

* 166435 245830: gap of unknown length

* 245830: contig of 79296 bp in length.

FEATURES

source

1. .245830
 /organism="Mus musculus"
 /mol_type="genomic DNA"
 /db_xref="taxon:10090"
 /chromosome="XNk"
 /clone="RP24-324C6"
 1. .1085
 /note="assembly_name:Contigs"
 1186. .11944
 /note="assembly_name:Contig8"
 12045. .30225
 /note="assembly_name:Contig9"
 30326. .63384
 /note="assembly_name:Contig10"
 63485. .117854
 /note="assembly_name:Contig11"
 117955. .166434
 /note="assembly_name:Contig12"
 166535. .245830
 /note="assembly_name:Contig13"

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

ORIGIN

Alignment Scores:

Pred. No.: 488 Length: 245830

Score: 58.00 Matches: 11

Percent Similarity: 70.59% Conservative: 1

Best Local Similarity: 64.71% Mismatches: 5

Query Match: 58.59% Indels: 0

DB: 2 Gaps: 0

US-10-799-747-116 (1-20) x AC140313 (1-245830)

QY 1 MetAlaAlaHisSerValLeuLeuThrProTyraLeu 17

Db 14529 ATGAAGACTTCACAAAGCTCTTCTATCTCTGTGGACCCCATAGTC 14479

RESULT 10

AC013568/c

LOCUS

DEFINITION Homo sapiens clone RP11-1B9, WORKING DRAFT SEQUENCE, 10 unordered

pieces.

ACCESSION

AC013568

VERSION

AC013568.3 GI:7107832

KEYWORDS

HTG; HTGS PHASE1; HTGS_DRAFT.

SOURCE

Homo sapiens (human)

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE

1 (bases 1 to 156471)

AUTHORS

Birren, B., Linton, L., Nusbaum, C., Lander, E., Allen, N., Anderson, M.,

Baldwin, J., Barna, N., Becker, R., Bozislavsky, L., Boukhgalter, B.,

Brown, A., Castle, A., Colangelo, M., Collins, S., Collymore, A.,

Cooke, P., DeArillano, K., Dewar, K., Domino, M., Donelan, L., Doyle, M.,

Ferreira, P., FitzHugh, W., Forrest, C., Funke, R., Gage, D.,

Galagan, J., Gardyna, S., Grant, G., Hagos, B., Heaford, A., Horton, L.,
 Howland, J. C., Johnson, R., Jones, C., Kann, L., Karatas, A., Klein, J.,
 Lehoczy, J., Lieu, C., Locke, K., Macdonald, P., Marquis, N.,
 McEwan, P., McGurk, A., McKernan, K., McLaughlin, J., Meldrim, J.,
 Morrow, J., Naylor, J., Norman, C. H., O'Connor, T., O'Donnell, P.,
 Peterson, K., Pollara, V., Riley, R., Roy, A., Santos, R., Severy, P.,
 Stange-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J.,
 Tesfaye, S., Tirrell, A., Vassiliev, H., Vo, A., Wheeler, J., Wu, X.,
 Wyman, D., Ye, W. J., Zimmer, A. and Zody, M.

Direct Submission

Submitted (13-NOV-1999) Whitehead Institute/MIT Center for Genome

Research, 320 Charles Street, Cambridge, MA 02141, USA

On Feb 28, 2000 this sequence version replaced gi:6514007.

All repeats were identified using RepeatMasker:

Smit, A.F.A. & Green, P. (1996-1997)

http://ftp.genome.washington.edu/RM/RepeatMasker.html

----- Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WIBR

Web site: http://www-seq.wi.mit.edu

Contact: sequence_submissions@genome.wi.mit.edu

----- Project Information

Center project name: L2468

Center clone name: L_B9

----- Summary Statistics

Sequencing vector: M13; M7815; 100% of reads

Chemistry: Dye-terminator Big Dye; 100% of reads

Assembly program: Phrap; version 0.960731

Consensus quality: 136271 bases at least Q40

Consensus quality: 149734 bases at least Q30

Consensus quality: 153894 bases at least Q20

Insert size: 170000; agarose-fp

Insert size: 155571; sum-of-contigs

Quality coverage: 5.4 in Q20 bases; agarose-fp

Quality coverage: 5.9 in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently
 * consists of 10 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

1 1916: contig of 1916 bp in length

* 1917 2016: gap of 100 bp

* 2017 4993: contig of 2977 bp in length

* 4994 5093: gap of 100 bp

* 5094 10286: contig of 5193 bp in length

* 10287 10386: gap of 100 bp

* 10387 15563: contig of 5177 bp in length

* 15564 15663: gap of 100 bp

* 15664 27659: contig of 11996 bp in length

* 27660 27759: gap of 100 bp

* 27760 41302: contig of 13543 bp in length

* 41303 41402: gap of 100 bp

* 41403 54013: contig of 12611 bp in length

* 54014 54113: gap of 100 bp

* 54114 75521: contig of 21408 bp in length

* 75522 75621: gap of 100 bp

* 75622 109229: contig of 33508 bp in length

* 109130 109229: gap of 100 bp

* 109230 156471: contig of 47242 bp in length.

FEATURES

source

1. .156471
 /organism="Homo sapiens"
 /mol_type="genomic DNA"
 /db_xref="taxon:9606"
 /clone="RP11-1B9"
 /clone_lib="RP11-1B9 Human Male BAC"
 1. .1916
 /note="assembly_fragment"
 2017. .4993
 misc_feature
 misc_feature

```

/misc_feature      /note="assembly_fragment"
5094. .10286
/note="assembly_fragment"
/misc_feature      10387. .15563
/note="assembly_fragment
clone end:T7
vector_side:right"
/misc_feature      15664. .27659
/note="assembly_fragment"
27760. .41302
/note="assembly_fragment"
/misc_feature      41403. .54013
/note="assembly_fragment"
/misc_feature      54114. .75521
/note="assembly_fragment"
75622. .109129
/note="assembly_fragment"
/misc_feature      109230. .156471
/note="assembly_fragment
clone end:SP6
vector_side:left"

ORIGIN
Alignment Scores:      444      Length:      156471
Pred. No.:      57.00      Matches:      10
Percent Similarity:      81.25%      Conservative:      3
Best Local Similarity:      62.50%      Mismatches:      3
Query Match:      57.58%      Indels:      0
DB:      2      Gaps:      0

US-10-799-747-116 (1-20) x AC013568 (1-156471)

QY      2 AlaAlaHisSerValLeuSerPheLeuLeuThrProTyraLeu 17
      ::::||||| ||| ||||| ||||| ||||| ||||| |||||
Db      146405 TCTAGTCATCCCATCTTTCCTTCTGAGCTGCATCTTACAGCTCG 146358

RESULT 11
AL358815/c
LOCUS      AL358815      192867 bp      DNA      linear      PRI 02-FEB-2001
DEFINITION      Human DNA sequence from clone RP11-52603 on chromosome 9, complete
sequence.
ACCESSION      AL358815
VERSION      AL358815.12      GI:12584727
KEYWORDS      HTG.
SOURCE      Homo sapiens (human)
ORGANISM      Homo sapiens
Bukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
Laird,G.
Direct Submission
Submitted (02-FEB-2001) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
Requests: clonerequest@sanger.ac.uk
On Jan 28, 2001 this sequence version replaced gi:12581015.
During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
corresponding to the overlapping clone, as we submit sequences with
only a small overlap as described above.
This sequence has been finished according to sequence map criteria
as follows. An attempt is made to resolve all sequencing problems,
such as compressions and repeats, but not necessarily within known
annotated repeat sequence elements. Where the sequence is
ambiguous, there is an annotation using the 'unsure' feature key.
The following abbreviations are used to associate primary accession
numbers given in the feature table with their source databases:
Em.; EMBL; Sw.; SWISSPROT; Tr.; TrEMBL; Wp.; WORMPEP; Information
on the WORMPEP database can be found at
http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
was generated from part of bacterial clone contigs of human

```

chromosome 9, constructed by the Sanger Centre Chromosome 9 Mapping Group. Further information can be found at <http://www.sanger.ac.uk/HGP/Chr9>
 RP11-52603 is from the library RPCI-11.2 constructed by the group of Pieter de Jong. For further details see <http://www.chori.org/bacpac/home.htm>
 VECTOR: pBACe3.6

This sequence is the entire insert of clone RP11-52603 The true left end of clone RP11-18A3 is at 135297 in this sequence.

FEATURES

```

source
1. .192867
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosomes="9"
/clone="RP11-52603"
/clone_lib="RPCI-11.2"
328. .681
/note="177 copies 2 mer aa 54% conserved"
repeat_region
2745. .2293
/note="L2 repeat: matches 2143. .2387 of consensus"
repeat_region
3071. .3179
/note="L1MD2 repeat: matches 5962. .6069 of consensus"
repeat_region
3188. .3408
/note="L2 repeat: matches 1908. .2140 of consensus"
repeat_region
3515. .3774
/note="AluSx repeat: matches 1. .291 of consensus"
repeat_region
3775. .3844
/note="AluSx repeat: matches 220. .290 of consensus"
repeat_region
4518. .4577
/note="L2 repeat: matches 1434. .1492 of consensus"
repeat_region
4554. .4736
/note="L2 repeat: matches 2242. .2419 of consensus"
repeat_region
4756. .4956
/note="MIR repeat: matches 2. .211 of consensus"
repeat_region
4957. .5256
/note="AluSx repeat: matches 1. .300 of consensus"
repeat_region
5257. .5309
/note="MIR repeat: matches 211. .261 of consensus"
repeat_region
7073. .7408
/note="MT1A1 repeat: matches 1. .365 of consensus"
repeat_region
7657. .7973
/note="AluJo repeat: matches 1. .286 of consensus"
repeat_region
7979. .8189
/note="MIR repeat: matches 47. .252 of consensus"
repeat_region
8329. .8463
/note="AluJo/FRAM repeat: matches 178. .311 of consensus"
repeat_region
10397. .10739
/note="L1MB5 repeat: matches 4107. .4444 of consensus"
repeat_region
10740. .11127
/note="L1MA2 repeat: matches 5685. .6079 of consensus"
repeat_region
11128. .11429
/note="AluSx repeat: matches 5. .305 of consensus"
repeat_region
11430. .11661
/note="L1MA2 repeat: matches 6079. .6308 of consensus"
repeat_region
11662. .13015
/note="L1MB5 repeat: matches 4444. .5403 of consensus"
repeat_region
12986. .13083
/note="L1MB5 repeat: matches 5994. .6096 of consensus"
repeat_region
13099. .13391
/note="AluSc repeat: matches 1. .290 of consensus"
repeat_region
13417. .13561
/note="AluJo/FRAM repeat: matches 154. .300 of consensus"
repeat_region
14426. .14737
/note="AluSx repeat: matches 1. .301 of consensus"
repeat_region
15741. .15784
/note="22 copies 2 mer aa 75% conserved"
repeat_region
17546. .17778
/note="MIR repeat: matches 16. .262 of consensus"
repeat_region
19580. .19879
/note="AluSg repeat: matches 1. .301 of consensus"
repeat_region
20116. .20396
/note="AluSg repeat: matches 1. .281 of consensus"

```



```

repeat_region 21132..21417
/notes="AluSq repeat: matches 5. .304 of consensus"
repeat_region 21828..21879
/notes="MIR repeat: matches 47. .97 of consensus"
repeat_region 22385..22689
/notes="AluSq repeat: matches 1. .308 of consensus"
repeat_region 22843..23149
/notes="AluSq repeat: matches 1. .308 of consensus"
repeat_region 23335..23431
/notes="MIR repeat: matches 158. .260 of consensus"
repeat_region 23380..23441
/notes="L2 repeat: matches 2648. .2709 of consensus"
repeat_region 23697..23837
/notes="FTAM C repeat: matches 1. .133 of consensus"
repeat_region 23850..24147
/notes="AluSq repeat: matches 1. .302 of consensus"
repeat_region 24429..24737
/notes="AluSq repeat: matches 1. .301 of consensus"
repeat_region 25297..25354
/notes="MIR repeat: matches 56. .120 of consensus"
misc_feature 26523..27318
/notes="CpG island"
/notes="evidence=not experimental"
repeat_region 27696..27925
/notes="MIR repeat: matches 6. .241 of consensus"
repeat_region 27923..28001
/notes="MER58 repeat: matches 6. .87 of consensus"
repeat_region 28069..28186
/notes="MIR repeat: matches 67. .182 of consensus"
repeat_region 28585..29048
/notes="LTR36 repeat: matches 107. .612 of consensus"
repeat_region 29049..29337
/notes="AluSq repeat: matches 1. .296 of consensus"
repeat_region 29338..29451
/notes="LTR36 repeat: matches 1. .107 of consensus"
repeat_region 29947..30256
/notes="AluSq repeat: matches 1. .309 of consensus"
repeat_region 30780..30953
/notes="L2 repeat: matches 2513. .2710 of consensus"
repeat_region 31789..31966
/notes="MIR repeat: matches 34. .221 of consensus"
repeat_region 31967..32242
/notes="AluSq repeat: matches 1. .276 of consensus"
repeat_region 32243..32257
/notes="MIR repeat: matches 221. .234 of consensus"
repeat_region 32405..32548
/notes="L2 repeat: matches 2374. .2509 of consensus"
repeat_region 32765..32967
/notes="MIR repeat: matches 20. .235 of consensus"
repeat_region 33210..33501
/notes="AluSq repeat: matches 21. .309 of consensus"
repeat_region 34265..34555
/notes="AluSq repeat: matches 1. .292 of consensus"
repeat_region 34621..34759
/notes="AluSq repeat: matches 1. .139 of consensus"
repeat_region 34877..35100
/notes="L2 repeat: matches 2065. .2332 of consensus"
repeat_region 35112..35449
/notes="LTR12 repeat: matches 12. .376 of consensus"
repeat_region 35450..35759
/notes="AluSq repeat: matches 1. .309 of consensus"
repeat_region 35760..35891
/notes="LTR12 repeat: matches 376. .501 of consensus"
repeat_region 36022..36391
/notes="LTR1A1 repeat: matches 1. .365 of consensus"
repeat_region 36553..36670
/notes="L2 repeat: matches 2582. .2699 of consensus"
repeat_region 36672..36837
/notes="FRAM repeat: matches 2. .176 of consensus"
repeat_region 38859..38936
/notes="MER47A repeat: matches 293. .366 of consensus"
repeat_region 38976..39054
/notes="MER47 repeat: matches 1. .79 of consensus"

```

```

repeat_region 39284..39594
/notes="AluSq repeat: matches 1. .309 of consensus"
repeat_region 39959..40181
/notes="MER58 repeat: matches 1. .229 of consensus"
repeat_region 40176..40253
/notes="MER58 repeat: matches 2214. .2420 of consensus"
repeat_region 40954..41068
/notes="MIR repeat: matches 117. .231 of consensus"
repeat_region 41524..41559
/notes="L2 repeat: matches 2711. .2746 of consensus"
repeat_region 41699..41792
/notes="L2 repeat: matches 2407. .2505 of consensus"
repeat_region 42767..43070
/notes="AluSq repeat: matches 1. .303 of consensus"
repeat_region 43268..43568
/notes="AluSq repeat: matches 1. .298 of consensus"
repeat_region 43854..44142
/notes="AluSq repeat: matches 1. .287 of consensus"
repeat_region 44450..44600
/notes="MIR repeat: matches 106. .262 of consensus"

```

Alignment Scores:

```

Pred. No.: 563 Length: 192867
Score: 57.00 Matches: 10
Percent Similarity: 81.25% Conservative: 3
Best Local Similarity: 62.50% Mismatches: 3
Query Match: 57.58% Indels: 0
DB: 9 Gaps: 0

```

US-10-799-747-116 (1-20) x AL359815 (1-192867)

QY 2 AlaAlaHisSerValLeuSerPheLeuLeuThrProTyrAlaLeu 17

Db 183796 TCTAGTCATTCCCATCTTTCCTTCGAGCTGACTCCTTACAGTCTG 183749

RESULT 12

AC024680/c

LOCUS Homo sapiens chromosome 8, clone RP11-347D13, complete sequence. 81715 bp DNA linear PRI 25-JUN-2002

AC024680

AC024680.10 GI:21553261

HTG. Homo sapiens (human)

SOURCE Homo sapiens

ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 81715)

Birren,B., Linton,L., Nusbaum,C. and Lander,E.

Homo sapiens chromosome 8, clone RP11-347D13

UNPUBLISHED

REFERENCE 2 (bases 1 to 81715)

Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,

Anderson,S., Baldwin,J., Barna,N., Bastien,V., Beda,F.,

Boguslavsky,L., Bouckgalter,B., Brown,A., Burkett,G.,

Campolano,A., Castie,A., Choepel,Y., Colangelo,M., Collins,S.,

Collumore,A., Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S.,

Dodge,S., Domino,M., Doyle,M., Ferreira,P., FitzHugh,W., Gage,D.,

Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,

Grand-pierre,N., Grant,G., Hagos,B., Hearford,A., Horton,L.,

Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,

Klein,U., Lacroque,K., Lamazares,R., Landers,T., Lehoczy,J.,

Levine,R., Liu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N.,

McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McPheeters,R.,

Meldrum,J., Meneus,L., Mihova,T., Miranda,C., Menga,V., Morrow,J.,

Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,

O'Neill,D., Olivari,I.M., Oliver,J., Peterson,K., Pierre,N.,

Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D.,

Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,

Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,

Tesfaye,S., Theodore,J., Tirrell,A., Travers,M., Trigilio,J.,

Vassiliev,H., Viel,R., Vo.A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,

Young,G., Zainoun,J., Zimmer,A. and Zody,M.

Direct Submission

TITLE


```

repeat_region      /rpt family="MER82"
complement(22487..22530)
/rpt family="L1MC4"
complement(22537..22845)
/rpt family="L1MC4a"
22859..22911
/rpt family=" (TCTA)n"
complement(22913..23034)
/rpt family="L1MC4"
23402..23422
/rpt family="AT_rich"
23642..23725
/rpt family="AT_rich"
complement(24308..24493)
/rpt family="L1ME"
26814..26776
/rpt family="MER103"
complement(27523..28419)

Alignment Scores:
Pred. No.:      323      Length:      81715
Score:          56.00      Matches:      9
Percent Similarity: 78.57%      Mismatches: 2
Best Local Similarity: 64.29%      Indels:      0
Query Match:     56.57%      Gaps:        0
DB:              9

US-10-799-747-116 (1-20) x AC024680 (1-81715)

QY 1 MetalAlaHisSerValLeuSerPheLeuLeuThrPro 14
Db 51079 ATGCCACAACTTCGATTGACGTTTATTCTCGACCCCA 51038

RESULT 13
AC091047/c
LOCUS      AC091047      165127 bp      DNA      linear
DEFINITION Homo sapiens chromosome 8, clone RP11-102F4, complete sequence.
ACCESSION AC091047
VERSION   AC091047.10  GI:23238050
KEYWORDS  HTG.
SOURCE    Homo sapiens (human)
ORGANISM  Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 165127)
AUTHORS   Birren,B., Nusbaum,C. and Lander,E.
TITLE     Homo sapiens chromosome 8, clone RP11-102F4
JOURNAL   Unpublished
PUBLISHED 2 (bases 1 to 165127)
AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,S.,
Barna,N., Bastien,V., Boguslavskiy,L., Boukhgalter,B., Brown,A.,
Camarata,J., Campopiano,A., Chang,J., Choepel,Y., Colangelo,M.,
Collins,S., Collymore,A., Cooke,P., DeArellano,K., Dewar,K.,
Diaz,J.S., Dodge,S., Faro,S., Ferreira,P., FitzGerald,M., Gage,D.,
Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
Grand-Pierre,N., Hagos,B., Heaford,A., Horton,L., Hulme,W.,
Iliev,I., Johnson,R., Jones,C., Karatas,A., Lander,E., Lander,E.,
Lamazzari,R., Landers,T., Lechoczy,J., Levine,R., Liu,G.,
Maclean,C., Macdonald,P., Marquis,N., Matthews,C., McCarthy,M.,
McEwan,P., McKernan,K., McPheeters,R., Meldrim,J., Meneus,L.,
Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Norbu,C.,
Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J.,
Peterson,K., Phunkhang,P., Pierre,N., Pollara,V., Raymond,C.,
Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P., Roman,J.,
Rosetti,M., Roy,A., Santos,R., Schauer,S., Schuback,R., Seaman,S.,
Severy,P., Sougnez,C., Spencer,B., Stange-Thomann,N., Talamas,J.,
Tesfaye,S., Theodorovic,N., Strauss,N., Subramanian,A., Talamas,J.,
Theodore,J., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,
Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
Direct Submission
Submitted (25-MAR-2001) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA

TITLE
JOURNAL
AUTHORS
REFERENCE

```

REFERENCE
AUTHORS

3 (bases 1 to 165127)
 Birren,B., Nusbaum,C., Lander,E., Ali,A., Allen,N., Anderson,S.,
 Barna,N., Bastien,V., Bloom,T., Boguslavskiy,L., Boukhgalter,B.,
 Camarata,J., Chang,J., Chazaro,B., Choepel,Y., Collymore,A.,
 Cook,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
 Faro,S., Ferreira,P., FitzGerald,M., Gage,D., Galagan,J.,
 Gardyna,S., Ginde,S., Graham,L., Grand-Pierre,N., Hagos,B.,
 Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Kamat,A.,
 Karatas,A., Kells,C., Landers,T., Levine,R., Lindblad-Toh,K.,
 Liu,G., Maclean,C., Macdonald,P., Major,J., Matthews,C.,
 McCarthy,M., Meldrim,J., Meneus,L., Mihova,T., Mlenga,V.,
 Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C., Norman,C.H.,
 O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J., Peterson,K.,
 Phunkhang,P., Pierre,N., Raymond,C., Retta,R., Rise,C., Rogov,P.,
 Roman,J., Roy,A., Schauer,S., Schuback,R., Seaman,S., Severy,P.,
 Smith,C., Spencer,B., Stange-Thomann,N., Stojanovic,N., Talamas,J.,
 Tesfaye,S., Theodore,J., Topham,K., Travers,M., Vassiliev,H.,
 Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Young,G., Zainoun,J.,
 Zembek,L., Zimmer,A. and Zody,M.

TITLE
JOURNAL

Submitted (20-AUG-2002) Whitehead Institute/MIT Center for Genome
 Research, 320 Charles Street, Cambridge, MA 02141, USA

REFERENCE
AUTHORS

4 (bases 1 to 165127)
 Birren,B., Nusbaum,C., Lander,E., Ali,A., Allen,N., Anderson,S.,
 Barna,N., Bastien,V., Bloom,T., Boguslavskiy,L., Boukhgalter,B.,
 Camarata,J., Chang,J., Chazaro,B., Choepel,Y., Collymore,A.,
 Cook,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
 Faro,S., Ferreira,P., FitzGerald,M., Gage,D., Galagan,J.,
 Gardyna,S., Ginde,S., Graham,L., Grand-Pierre,N., Hagos,B.,
 Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Kamat,A.,
 Karatas,A., Kells,C., Landers,T., Levine,R., Lindblad-Toh,K.,
 Liu,G., Maclean,C., Macdonald,P., Major,J., Matthews,C.,
 McCarthy,M., Meldrim,J., Meneus,L., Mihova,T., Mlenga,V.,
 Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C., Norman,C.H.,
 O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J., Peterson,K.,
 Phunkhang,P., Pierre,N., Raymond,C., Retta,R., Rise,C., Rogov,P.,
 Roman,J., Roy,A., Schauer,S., Schuback,R., Seaman,S., Severy,P.,
 Smith,C., Spencer,B., Stange-Thomann,N., Stojanovic,N., Talamas,J.,
 Tesfaye,S., Theodore,J., Topham,K., Travers,M., Vassiliev,H.,
 Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Young,G., Zainoun,J.,
 Zembek,L., Zimmer,A. and Zody,M.

TITLE
JOURNAL

Submitted (20-SEP-2002) Whitehead Institute/MIT Center for Genome
 Research, 320 Charles Street, Cambridge, MA 02141, USA

COMMENT

On Sep 20, 2002 this sequence version replaced gi:22325339.
 All repeats were identified using RepeatMasker:
 Smit, A.F.A. & Green, P. (1996-1997)
 http://ftp.genome.washington.edu/RM/RepeatMasker.html

TITLE
JOURNAL

Center: Whitehead Institute/ MIT Center for Genome Research
 Center code: WIBR

Web site: <http://www-seq.wi.mit.edu>

Contact: sequence_submissions@genome.wi.mit.edu

Center project name: Li0484

Center clone name: 102_F_4

FEATURES

source

Location/Qualifiers

1..165127
 /organism="Homo sapiens"
 /mol_type="genomic DNA"
 /db_xref="taxon:9606"
 /maposome="8"
 /map="8"
 /clone="RP11-102F4"
 /clone_lib="RPC1-11 Human Male BAC"
 66..176
 /rpt family="L1MB5"
 complement(1210..1513)
 /rpt family="L2"
 1797..1803
 /note="<30 qual SNGL region"
 1821..1825

repeat_region

repeat_region

unsure

unsure

```
repeat_region /note="<30 qual SNGL region"
repeat_region 2411..2532
repeat_region /rpt_family="FRAM"
repeat_region 2534..2556
repeat_region /rpt_family="CAAA)n"
repeat_region 2805..3112
repeat_region /rpt_family="AluSg"
repeat_region complement(3673..3776)
repeat_region /rpt_family="L1MC4"
repeat_region complement(3808..4394)
repeat_region /rpt_family="L1MC4"
repeat_region 4395..5019
repeat_region /rpt_family="MER82"
repeat_region complement(5020..5063)
repeat_region /rpt_family="L1MC4"
repeat_region complement(5070..5378)
repeat_region /rpt_family="L1MC4a"
repeat_region 5392..5444
repeat_region /rpt_family="(TCTA)n"
repeat_region complement(5446..5567)
repeat_region /rpt_family="L1MC4"
repeat_region 5935..5955
repeat_region /rpt_family="AT_rich"
repeat_region 6175..6258
repeat_region /rpt_family="AT_rich"
repeat_region complement(6841..7026)
repeat_region /rpt_family="L1ME"
repeat_region 9147..9309
repeat_region /rpt_family="MER103"
repeat_region complement(10056..10952)
repeat_region /rpt_family="L1PAL0"
repeat_region complement(12015..12381)
repeat_region /rpt_family="THEIC"
repeat_region 12402..12570
repeat_region /rpt_family="WIR"
repeat_region complement(12765..12978)
repeat_region /rpt_family="WIR"
repeat_region complement(13005..13247)
repeat_region /rpt_family="AluSc"
repeat_region complement(13285..13325)
repeat_region /rpt_family="AluSc"
repeat_region complement(14540..14604)
repeat_region /rpt_family="WAD81"
repeat_region 14605..14902
repeat_region /rpt_family="AluSx"
repeat_region complement(14903..14928)
repeat_region /rpt_family="WAD81"
repeat_region 15710..15892
repeat_region /rpt_family="WIR"
repeat_region complement(16294..16431)
repeat_region /rpt_family="L1MC5"
repeat_region complement(16432..16740)
repeat_region /rpt_family="AluY"
repeat_region complement(16741..17889)
repeat_region /rpt_family="L1MC5"
repeat_region complement(18258..18416)
repeat_region /rpt_family="WIR"
repeat_region 18631..18692
repeat_region /rpt_family="(TCCA)n"
repeat_region complement(18695..18804)
repeat_region /rpt_family="L2"
repeat_region 18867..18939
repeat_region /rpt_family="WIR"
repeat_region 19104..19407
repeat_region /rpt_family="AluSx"
repeat_region 20410..20443
repeat_region /rpt_family="(CATA)n"
repeat_region 20645..20850
repeat_region /rpt_family="WIR"
repeat_region complement(21039..21092)
repeat_region /rpt_family="WIR"
repeat_region 21832..22130
repeat_region /rpt_family="AluSx"

repeat_region 22358..22661
repeat_region /rpt_family="AluJo"
repeat_region 23062..23127
repeat_region /rpt_family="(CA)n"
repeat_region 23297..23425
repeat_region /rpt_family="MIR"
repeat_region 23516..23604
repeat_region /rpt_family="MIR"
repeat_region complement(23704..24014)
repeat_region /rpt_family="AluSx"
repeat_region 24368..24523
repeat_region /rpt_family="MIR3"
repeat_region complement(24564..24647)
repeat_region /rpt_family="MLT1J"
repeat_region complement(24787..24892)
repeat_region /rpt_family="MLT1J"

Alignment Scores:
Pred. No.: 717 Length: 165127
Score: 56.00 Matches: 9
Percent Similarity: 78.57% Conservative: 2
Best Local Similarity: 64.29% Mismatches: 3
Query Match: 56.57% Indels: 0
DB: 9 Gaps: 0

US-10-799-747-116 (1-20) x AC091047 (1-165127)
Qy 1 MetAlaAlaHisSerValLeuSerPheLeuLeuTriThrPro 14
Db 33611 ATGCCACAACTCGATTGACGTTTTATTCTGACCCCA 33570

RESULT 14
AC140381 233717 bp DNA linear HTG 23-FEB-2003
LOCUS Mus musculus chromosome UNK clone RP23-176M4, WORKING DRAFT
DEFINITION SEQUENCE, 18 unordered pieces.
ACCESSION AC140381
VERSION AC140381.1 GI:28475640
KEYWORDS HTG; HTGS PHASE1; HTGS DRAFT; HTGS_FULLTOP.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
REFERENCE 1 (bases 1 to 233717)
AUTHORS McPherson,J.D. and Waterston,R.H.
TITLE The sequence of Mus musculus clone
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 233717)
AUTHORS McPherson,J.D. and Waterston,R.H.
TITLE Direct Submission
JOURNAL Submitted (23-FEB-2003) Genome Sequencing Center, 4444 Forest Park
Parkway, St. Louis, MO 63108, USA

COMMENT
----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc/index.shtml
Contact: submissions@watson.wustl.edu
----- Project Information -----
Center project name: M_BA0176M04
----- Summary Statistics -----
Sequencing vector: M13, 0%
Sequencing vector: plasmid, 100%
Chemistry: Dye-primer ET; 0% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 226925 bases at least Q40
Consensus quality: 228328 bases at least Q30
Consensus quality: 229211 bases at least Q20
* NOTE: This is a 'working draft' sequence. It currently
* consists of 18 contigs. The true order of the pieces
```

* is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

```

1 1463: contig of 1463 bp in length
* 1464 1563: gap of unknown length
* 1564 3655: contig of 2092 bp in length
* 3656 3755: gap of unknown length
* 3756 7668: contig of 3913 bp in length
* 7669 7769: gap of unknown length
* 7769 11494: contig of 3726 bp in length
* 11495 11594: gap of unknown length
* 11595 13492: contig of 1898 bp in length
* 13493 13592: gap of unknown length
* 13593 17610: contig of 4018 bp in length
* 17611 17710: gap of unknown length
* 17711 2310: contig of 5500 bp in length
* 2311 23310: gap of unknown length
* 23311 27592: contig of 4282 bp in length
* 27593 27692: gap of unknown length
* 27693 34114: contig of 6422 bp in length
* 34115 34214: gap of unknown length
* 34215 45449: contig of 11235 bp in length
* 45450 45499: gap of unknown length
* 45500 56906: contig of 11357 bp in length
* 56907 57006: gap of unknown length
* 57007 67070: contig of 10064 bp in length
* 67071 67170: gap of unknown length
* 67171 80062: contig of 12892 bp in length
* 80063 80162: gap of unknown length
* 80163 104523: contig of 24361 bp in length
* 104524 104623: gap of unknown length
* 104624 126747: contig of 22124 bp in length
* 126748 126847: gap of unknown length
* 126848 156641: contig of 29794 bp in length
* 156642 156741: gap of unknown length
* 156742 188399: contig of 31558 bp in length
* 188399 188300: gap of unknown length
* 188400 233717: contig of 45318 bp in length.

```

FEATURES

source

```

1..233717
/organism="Mus musculus"
/mol_type="genomic DNA"
/db_xref="taxon:10090"
/chromosomes="UNK"
/clone="RP23-176M4"

```

```

misc_feature 1..1463
/feature="assembly_name:Contig27"
misc_feature 1564..3655
/feature="assembly_name:Contig28"
misc_feature 3756..7668
/feature="assembly_name:Contig30"
misc_feature 7769..11494
/feature="assembly_name:Contig31"
misc_feature 11595..13492
/feature="assembly_name:Contig32"
misc_feature 13593..17610
/feature="assembly_name:Contig33"
misc_feature 17711..23210
/feature="assembly_name:Contig34"
misc_feature 23311..27592
/feature="assembly_name:Contig35"
misc_feature 27693..34114
/feature="assembly_name:Contig36"
misc_feature 34215..45449
/feature="assembly_name:Contig37"
misc_feature 45500..56906
/feature="assembly_name:Contig38"
misc_feature 57007..67070
/feature="assembly_name:Contig39"
misc_feature 67171..80062

```

```

/misc_feature 80163..104523
/feature="assembly_name:Contig41"
/misc_feature 104624..126747
/feature="assembly_name:Contig42"
/misc_feature 126848..156641
/feature="assembly_name:Contig43"
/misc_feature 156742..188299
/feature="assembly_name:Contig44"
/misc_feature 188400..233717
/feature="assembly_name:Contig45"

```

ORIGIN

```

Alignment Scores:
Pred. No.: 1.06e+03 Length: 233717
Score: 56.00 Matches: 9
Percent Similarity: 77.78% Conservative: 5
Best Local Similarity: 50.00% Mismatches: 4
Query Match: 56.57% Indels: 0
DB: 2 Gaps: 0

```

US-10-799-747-116 (1-20) x AC140381 (1-233717)

```

QY 2 AlaAlaHisSerValLeuSerPheLeuLeuTThrProTyAlaLeuLysSer 19
Db 126482 TCTGCCACTCTCTGTTTCTCCCATTTGGAGACCCCTACAGTCTGCTTCN 126535

```

RESULT 15

```

LOCUS AL137138 98558 bp DNA linear PRI 31-MAY-2001
DEFINITION Human DNA sequence from clone RP11-117113 on chromosome 13,
complete sequence.
ACCESSION AL137138
VERSION AL137138.13 GI:14280387
KEYWORDS HTG.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 98558)
AUTHORS Tromans, A.
DIRECT SUBMISSION
TITLE Submitted (31-MAY-2001) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
requests: clonerequest@sanger.ac.uk
COMMENT On Jun 3, 2001 this sequence version replaced gi:13785011.
During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
corresponding to the overlapping clone, as we submit sequences with
only a small overlap as described above.
This sequence was finished as follows unless otherwise noted: all
regions were either double-stranded or sequenced with an alternate
chemistry or covered by high quality data (i.e., phred quality >=
30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by at least
one plasmid subclone or more than one M13 subclone; and the
assembly was confirmed by restriction digest. The following
abbreviations are used to associate primary accession numbers given
in the feature table with their source databases: Em:, EMBL; Sw:,
SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information on the WORMPEP
database can be found at
http://www.sanger.ac.uk/Projects/C_elegans/wormpep
This sequence
was generated from part of bacterial clone contigs of human
chromosome 13, constructed by the Sanger Centre Chromosome 13
Mapping Group. Further information can be found at
http://www.sanger.ac.uk/HGP/Chr13
RP11-117113 is from the library RPC1-11.1 constructed by the group
of Pieter de Jong. For further details see
http://www.chori.org/bacpac/home.htm
VECTOR: pBAC3.6
IMPORTANT: This sequence is not the entire insert of clone

```

RP11-117113 It may be shorter because we sequence overlapping sections only once, except for a 100 base overlap.
The true left end of clone RP11-117113 is at 1 in this sequence.
The true left end of clone RP11-366017 is at 98459 in this sequence.

FEATURES

Source Location/Qualifiers

```

1..98558
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosomes="13"
/clone="RP11-117113"
/clone_lib="RPCI-11.1"
repeat_region
1380..1587
/notes="MER59A repeat: matches 6. .224 of consensus"
repeat_region
4303..4506
/notes="L1MA9 repeat: matches 6070. .6270 of consensus"
repeat_region
4536..4844
/notes="AluSg repeat: matches 1. .305 of consensus"
repeat_region
5066..5362
/notes="AluX repeat: matches 1. .298 of consensus"
repeat_region
5544..5674
/notes="L2 repeat: matches 2619. .2748 of consensus"
repeat_region
7241..8826
/notes="L1PB1 repeat: matches 4551. .6148 of consensus"
repeat_region
8828..9507
/notes="L1PB1 repeat: matches 2438. .2734 of consensus"
repeat_region
9596..10327
/notes="L1PBa repeat: matches -237. .1442 of consensus"
repeat_region
10350..11388
/notes="L1PBa repeat: matches -1537. -.412 of consensus"
misc_feature
12251..12425
/notes="match: STS: Em:G15057"
repeat_region
12414..12652
/notes="MER57-internal repeat: matches 7155. .7387 of consensus"
repeat_region
12876..13029
/notes="MLT1H repeat: matches 178. .340 of consensus"
repeat_region
13220..13577
/notes="MLT1I repeat: matches 43. .410 of consensus"
repeat_region
13578..13658
/notes="L2 repeat: matches 2420. .2496 of consensus"
repeat_region
13659..14025
/notes="THE1C repeat: matches 1. .369 of consensus"
repeat_region
14026..14254
/notes="L2 repeat: matches 2201. .2420 of consensus"
repeat_region
14255..14532
/notes="AluJo repeat: matches 1. .280 of consensus"
repeat_region
14533..14844
/notes="L2 repeat: matches 1862. .2201 of consensus"
repeat_region
15852..16106
/notes="L2 repeat: matches 2201. .2456 of consensus"
repeat_region
16209..16274
/notes="L2 repeat: matches 2133. .2202 of consensus"
repeat_region
16270..16309
/notes="L2 repeat: matches 2464. .2503 of consensus"
repeat_region
16351..16819
/notes="MER41B repeat: matches 130. .596 of consensus"
repeat_region
17215..17470
/notes="L2 repeat: matches 2438. .2701 of consensus"
repeat_region
17722..17851
/notes="MIR repeat: matches 61. .191 of consensus"
repeat_region
18490..18521
/notes="16 copies 2 mer tg 93% conserved"
repeat_region
19327..19402
/notes="MLT1F repeat: matches 89. .161 of consensus"
repeat_region
19430..19465
/notes="18 copies 2 mer tc 86% conserved"
repeat_region
19536..19782
/notes="MLT1F repeat: matches 244. .546 of consensus"
repeat_region
20422..20618
/notes="L2 repeat: matches 2554. .2747 of consensus"
repeat_region
21524..21645

```

```

/notes="AluSp/q repeat: matches 187. .303 of consensus"
21647..21868
/notes="AluX repeat: matches 1. .222 of consensus"
22326..22377
/notes="MIR repeat: matches 47. .97 of consensus"
24000..24027
/notes="14 copies 2 mer tg 89% conserved"
25338..25379
/notes="MIR repeat: matches 58. .99 of consensus"
25652..25812
/notes="MER44A repeat: matches 5. .170 of consensus"
25813..26530
/notes="trigger3(Golem) repeat: matches 2277. .3028 of consensus"
26524..26594
/notes="trigger3(Golem) repeat: matches 36. .106 of consensus"
26655..26806
/notes="MER44A repeat: matches 166. .329 of consensus"
26817..27149
/notes="MER46C repeat: matches 3. .338 of consensus"
27514..28149
/notes="LTR5 repeat: matches 1. .623 of consensus"
complement(27967..28529)
/notes="match: STS: Em:G42855"
28229..28529
/notes="LTR5 repeat: matches 675. .969 of consensus"
29151..29504
/notes="THE1A repeat: matches 1. .354 of consensus"
29676..29814
/notes="L2 repeat: matches 2597. .2750 of consensus"
31144..31774
/notes="L1ME1 repeat: matches 5460. .6156 of consensus"
31775..31911
/notes="AluJo repeat: matches 1. .85 of consensus"
31912..32186
/notes="AluSc repeat: matches 1. .281 of consensus"
32187..32371
/notes="AluJo repeat: matches 85. .288 of consensus"
32372..32588
/notes="L1ME1 repeat: matches 5254. .5460 of consensus"
33053..33329
/notes="AluSg repeat: matches 1. .276 of consensus"
33807..33871
/notes="L2 repeat: matches 2622. .2687 of consensus"
34562..34746
/notes="MER5A repeat: matches 1. .188 of consensus"
35475..35635
/notes="MIR repeat: matches 78. .230 of consensus"
35840..36257
/notes="L2 repeat: matches 1450. .1892 of consensus"
38733..39038
/notes="MLT1J repeat: matches 142. .468 of consensus"
39053..39147
/notes="MER86 repeat: matches 88. .182 of consensus"
39567..39723
/notes="L1MEC repeat: matches 1818. .1984 of consensus"
40423..40607
/notes="L1MA repeat: matches 2680. .2886 of consensus"
40608..41117
/notes="LTR40a repeat: matches 1. .519 of consensus"
41118..41187
/notes="L1MA repeat: matches 2886. .2950 of consensus"
41302..41405
/notes="52 copies 2 mer aa 60% conserved"
42443..42709
/notes="L2 repeat: matches 2457. .2710 of consensus"
43036..43484
/notes="MLT1C repeat: matches 7. .466 of consensus"
44299..44582
/notes="AluX repeat: matches 1. .296 of consensus"
46846..46936
/notes="MER91A repeat: matches 17. .113 of consensus"

```

```

repeat_region 48664..49299
/note="MER4D repeat: matches 362. .981 of consensus"
repeat_region 49298..49537
/note="MER49 repeat: matches 2. .241 of consensus"
repeat_region 49608..49961
/note="LTR37A repeat: matches 91. .421 of consensus"
repeat_region 49987..50012
/note="13 copies 2 mer gt 100% conserved"
repeat_region 51260..51623
/note="THE1B repeat: matches 1. .364 of consensus"

```

Alignment Scores:

Pred. No.:	607	Length:	98558
Score:	55.00	Matches:	9
Percent Similarity:	100.00%	Conservative:	5
Best Local Similarity:	64.29%	Mismatches:	0
Query Match:	55.56%	Indels:	0
DB:	9	Gaps:	0

US-10-799-747-116 (1-20) x AL137138 (1-98558)

QY 2 AlalahisservalleuSerPheLeuLeuTrpThrProTyr 15
 Db 25528 GCAAGCCACACAGTCATCTCTTCGTATATACCTTC 25487

Search completed: July 21, 2004, 02:13:58
 Job time : 1837 secs

This Page Blank (uspto)

GenCore version 5.1.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: July 21, 2004, 00:02:30 ; Search time 348 Seconds
(without alignments)
244.149 Million cell updates/sec

Title: US-10-799-747-116
Perfect score: 99
Sequence: 1 MAHVSFLWTPYALKSX 20

Scoring table:
BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 3373863 seqs, 2124099041 residues

Total number of hits satisfying chosen parameters: 6747726

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:

-O=/cpn2.1/USPTO.spool_p/US10799747/runat_19072004.161415.21434/app.query.fasta_1.199
-DB=N_Geneseq_29Jan04 -QFMT=fastap -SUFFIX=ring -MINMATCH=0.1 -LOOPEXT=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOCALLIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000
-USER=US10799747 @CGN 1 1 708 @runat_19072004.161415.21434 -NCPU=6 -ICPU=3
-NO WMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPELOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : N_Geneseq_29Jan04.*

- 1: Geneseqn1980s.*
- 2: Geneseqn1990s.*
- 3: Geneseqn2000s.*
- 4: Geneseqn2001as.*
- 5: Geneseqn2001bs.*
- 6: Geneseqn2002s.*
- 7: Geneseqn2003as.*
- 8: Geneseqn2003bs.*
- 9: Geneseqn2003cs.*
- 10: Geneseqn2004s.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	98	99.0	1434	2 AAX37452	Human sec
2	98	99.0	1434	7 ADA39771	Human sec
3	98	99.0	1434	7 ACC50424	Human dig
4	56	56.6	8417	4 AAK89495	Human gen
5	56	56.6	8417	4 AAS32681	Human gen
6	53	53.5	5760	4 AAS32472	Human gen
7	53	53.5	5763	4 AAS32471	Human gen
8	52	52.5	2877	7 ABZ36220	Human sec

C	9	52	52.5	36785	4	AAX82208	Human imm
C	10	51	51.5	400	5	AAF64369	Novel hum
C	11	51	51.5	505	4	AAH03292	Human cDN
C	12	51	51.5	708	5	AAF64473	Novel hum
C	13	51	51.5	1544	2	AAT86839	cDNA enco
C	14	51	51.5	1568	4	AAH15239	Human cDN
C	15	51	51.5	2500	4	AAH77480	Mandrill
C	16	51	51.5	9641	4	AAH77486	Mandrill
C	17	50	50.5	620	7	ABT40513	Toxicity
C	18	50	50.5	620	9	ADB50026	Primary r
C	19	49	49.5	6708	9	ADC56730	Thale cre
C	20	49	49.5	6708	9	ADC56727	Open read
C	21	49	49.5	7149	9	ADC56728	Thale cre
C	22	49	49.5	11000	9	ADC56726	Thale cre
C	23	49	49.5	32768	2	AAX13065	Enterococ
C	24	49	49.5	32768	6	ABS98860	Enterococ
C	25	48	48.5	207	7	ABX88609	Corn ear-
C	26	48	48.5	290	2	AAV88159	EST clone
C	27	48	48.5	375	4	AAS39268	Novel hum
C	28	48	48.5	404	3	AAF11614	Aspergill
C	29	48	48.5	481	6	ABN64079	Human can
C	30	48	48.5	485	5	ABV45709	Human pro
C	31	48	48.5	495	6	ABQ57387	Human col
C	32	48	48.5	612	6	ABY07338	Human lun
C	33	48	48.5	648	5	ABV15807	Human pro
C	34	48	48.5	675	5	ABV45609	Human pro
C	35	48	48.5	804	4	AAH32323	Human Olf
C	36	48	48.5	939	7	ABZ77940	Human G p
C	37	48	48.5	940	9	ADC86574	Human GPC
C	38	48	48.5	960	4	AAH12359	Human G-p
C	39	48	48.5	960	5	AA542282	Human cDN
C	40	48	48.5	960	6	ABZ43079	Human GPC
C	41	48	48.5	960	6	ABK68434	Human DNA
C	42	48	48.5	960	6	ABK37568	DNA encod
C	43	48	48.5	961	6	ABD28710	Human G-p
C	44	48	48.5	970	6	ABS58783	Human G-p
C	45	48	48.5	970	6	AAD28709	Human G-p

ALIGNMENTS

RESULT 1
AAX37452
ID AAX37452 standard; cDNA; 1434 BP.
XX
AC AAX37452;
XX
DT 06-JUL-1999 (first entry)
XX
DE Human secreted protein cDNA fragment containing gene 2.
XX
KW Human; secreted protein; treatment; prevention; protein therapy; AIDS;
KW gene therapy; diagnosis; cancer; tumour; neurodegenerative disorder;
KW developmental abnormality; fetal deficiency; blood disorder; leukemia;
KW immune system disease; autoimmune disease; hepatic disease; lymphoma;
KW renal disease; inflammation; allergy; Alzheimer's disease; schizophrenia;
KW cognitive disorder; prostate disease; skeletal; cardiac; muscle disorder;
KW pulmonary disorder; transplant rejection; osteoclast; osteoporosis;
KW arthritis; malignancy; digestive; endocrine; infection; ss.
XX
OS Homo sapiens.

Result No.

Score

Query Match

Length

ID

Description

PR 02-OCT-1997; 97US-0060843P.
 PR 02-OCT-1997; 97US-0060862P.
 PR 02-OCT-1997; 97US-0060866P.
 PR 02-OCT-1997; 97US-0060874P.
 PR 02-OCT-1997; 97US-0060880P.
 PR 02-OCT-1997; 97US-0060884P.
 XX (HUMA-) HUMAN GENOME SCI INC.
 PA
 XX Duan DR, Florence KA, Rosen CA, Ruben SM, Greene JM, Young P;
 PI Ferrie AM, Yu G, Janat F, Ni J, Carter KC, Endress GA, Feng P;
 PI Lafleur DW, Shi Y;
 XX WPI; 1999-264022/22.
 DR P-PSDB; AAY07853.
 DR
 XX New isolated human genes and the secreted polypeptides they encode.
 PT
 XX Claim 1a; Page 228; 368pp; English.
 PS
 CC This invention describes novel isolated human genes and the secreted
 CC proteins they encode. The products of the invention are useful for
 CC preventing, treating or ameliorating medical conditions, e.g. by protein
 CC or gene therapy. Also pathological conditions can be diagnosed by
 CC determining the amount of the new polypeptides in a sample or by
 CC determining the presence of mutations in the new polynucleotides.
 CC Specific uses are described for each of the 101 polynucleotides, based on
 CC which tissues they are most highly expressed in, and include developing
 CC products for the diagnosis or treatment of cancer, tumours,
 CC neurodegenerative disorders, developmental abnormalities and fetal
 CC deficiencies, blood disorders, leukemias, diseases of the immune system,
 CC autoimmune diseases, hepatic and renal disease, lymphomas, inflammation,
 CC allergies, Alzheimer's and cognitive disorders, schizophrenia, prostate
 CC disease, skeletal or cardiac muscle disorders, pulmonary disorders,
 CC transplant rejection, disorders involving osteoclasts such as
 CC osteoporosis, arthritis or malignancies, digestive/endocrine disorders,
 CC infections and AIDS. The human secreted proteins of the invention are
 CC represented in AAY07852-Y07993 and the encoding nucleic acids are
 CC represented in AAX37451-X37552
 XX
 SQ Sequence 1434 BP; 480 A; 204 C; 250 G; 495 T; 0 U; 5 Other;
 Alignment Scores:
 Pred. No.: 7.13e-07 Length: 1434
 Score: 98.00 Matches: 19
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 98.99% Indels: 0
 DB: 2 Gaps: 0
 US-10-799-747-116 (1-20) x AAX37452 (1-1434)
 QY 1 MetAlaAlaHisSerValLeuSerPheLeuLeuThrProTyrAlaLeuLysSer 19
 Db 507 ATGGCAGCCCATTCAGCTTTCAGTTTCTCTCTGGACACCTTATGCTCTGAAATCA 563
 RESULT 2
 ADA39771
 ID ADA39771 standard; cDNA; 1434 BP.
 XX
 AC ADA39771;
 XX
 XX 20-NOV-2003 (first entry)
 DT
 XX Human secreted protein encoding cDNA.
 DE
 XX Human; secreted protein; cancer; hyperproliferative disorder;
 KW rheumatoid arthritis; autoimmune disorder; haematopoietic disorder;
 KW anaemia; allergic reaction; asthma; cardiovascular disorder;
 KW wound healing; cytostatic; immunosuppressive; neutropenic; neuroprotective;
 KW antiviral; antiallergic; hepatotropic; antidiabetic; antiinflammatory;
 KW vulnary; cardiant; gene therapy; ss.

OS Homo sapiens.
 XX
 PN WO2002102993-A2.
 XX
 PD 27-DEC-2002.
 XX
 PF 19-MAR-2002; 2002WO-US008123.
 XX
 XX 21-MAR-2001; 2001US-0277340P.
 PR 19-JUL-2001; 2001US-0306171P.
 PR 13-NOV-2001; 2001US-0331287P.
 XX
 PA (HUMA-) HUMAN GENOME SCI INC.
 XX
 XX Rosen CA, Ruben SM;
 PI
 XX WPI; 2003-175238/17.
 DR
 XX New human secreted proteins and nucleic acid molecules, useful for
 PT preparing a diagnostic or pharmaceutical composition for diagnosing,
 PT preventing or treating cancer or other hyperproliferative disorder,
 PT asthma, allergies or AIDS.
 XX
 PS Claim 9; SEQ ID NO 153; 3205pp; English.
 XX
 CC The invention relates to novel genes ADA39629-ADA40565 and proteins
 CC ADA40566-ADA41501 for human secreted proteins, useful for preventing,
 CC treating or ameliorating medical conditions e.g. by protein or gene
 CC therapy. The polypeptides, nucleic acid molecules, antibodies or their
 CC fragments, and agonists or antagonists that bind to the polypeptide are
 CC useful for preparing a diagnostic or pharmaceutical composition for
 CC diagnosing or treating cancer or other hyperproliferative disorder. The
 CC polypeptides and nucleic acid molecules are also useful for detecting,
 CC preventing, diagnosing, prognosticating, treating or ameliorating cancer
 CC or other hyperproliferative disorders including neoplasms, autoimmune
 CC disorders (e.g. diabetes, rheumatoid arthritis, systemic lupus
 CC erythematosus, multiple sclerosis, autoimmune thyroiditis or haemolytic
 CC anaemia), haematopoietic or haematological disorders (e.g. anaemia,
 CC thrombocytopenia), allergic reactions including asthma or eczema,
 CC inflammatory disorders (e.g. ischaemia-reperfusion injury, inflammatory
 CC bowel disease or Crohn's disease), neurodegenerative disorders (e.g.
 CC Alzheimer's disease or Parkinson's disease), cardiovascular disorders
 CC (e.g. atherosclerosis, myocarditis), infectious diseases (bacterial,
 CC fungal or viral infections including HIV/AIDS), or wound healing and
 CC disorders of epithelial cell proliferation. The nucleic acids are also
 CC useful for chromosome identification, radiation hybrid mapping or long-
 CC range restriction mapping, as molecular weight markers, or as
 CC hybridization or diagnostic probes. The polypeptides and antibodies are
 CC useful for providing immunological probes for differential identification
 CC of the tissues immunohistochemistry assays. Note: The sequence data for
 CC this patent did not form part of the printed specification, but was
 CC obtained in electronic format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences.
 XX
 SQ Sequence 1434 BP; 480 A; 203 C; 250 G; 496 T; 0 U; 5 Other;
 Alignment Scores:
 Pred. No.: 7.13e-07 Length: 1434
 Score: 98.00 Matches: 19
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 98.99% Indels: 0
 DB: 7 Gaps: 0
 US-10-799-747-116 (1-20) x ADA39771 (1-1434)
 QY 1 MetAlaAlaHisSerValLeuSerPheLeuLeuThrProTyrAlaLeuLysSer 19
 Db 507 ATGGCAGCCCATTCAGCTTTCAGTTTCTCTCTGGACACCTTATGCTCTGAAATCA 563
 RESULT 3
 ACC50424
 ID ACC50424 standard; cDNA; 1434 BP.

PR 08-SEP-2000; 2000US-0232080P.
PR 08-SEP-2000; 2000US-0232081P.
PR 12-SEP-2000; 2000US-0231968P.
PR 14-SEP-2000; 2000US-0232397P.
PR 14-SEP-2000; 2000US-0232398P.
PR 14-SEP-2000; 2000US-0232399P.
PR 14-SEP-2000; 2000US-0232400P.
PR 14-SEP-2000; 2000US-0232401P.
PR 14-SEP-2000; 2000US-0233063P.
PR 14-SEP-2000; 2000US-0233064P.
PR 14-SEP-2000; 2000US-0233065P.
PR 21-SEP-2000; 2000US-0234223P.
PR 21-SEP-2000; 2000US-0234274P.
PR 25-SEP-2000; 2000US-0234997P.
PR 25-SEP-2000; 2000US-0234998P.
PR 26-SEP-2000; 2000US-0235484P.
PR 27-SEP-2000; 2000US-0235834P.
PR 27-SEP-2000; 2000US-0235836P.
PR 29-SEP-2000; 2000US-0236327P.
PR 29-SEP-2000; 2000US-0236367P.
PR 29-SEP-2000; 2000US-0236368P.
PR 29-SEP-2000; 2000US-0236369P.
PR 29-SEP-2000; 2000US-0236370P.
PR 02-OCT-2000; 2000US-0236802P.
PR 02-OCT-2000; 2000US-0237037P.
PR 02-OCT-2000; 2000US-0237038P.
PR 02-OCT-2000; 2000US-0237039P.
PR 02-OCT-2000; 2000US-0237040P.
PR 13-OCT-2000; 2000US-0239335P.
PR 13-OCT-2000; 2000US-0239337P.
PR 20-OCT-2000; 2000US-0240960P.
PR 20-OCT-2000; 2000US-0241221P.
PR 20-OCT-2000; 2000US-0241785P.
PR 20-OCT-2000; 2000US-0241786P.
PR 20-OCT-2000; 2000US-0241787P.
PR 20-OCT-2000; 2000US-0241808P.
PR 20-OCT-2000; 2000US-0241809P.
PR 20-OCT-2000; 2000US-0241809P.
PR 01-NOV-2000; 2000US-0241826P.
PR 01-NOV-2000; 2000US-0244617P.
PR 08-NOV-2000; 2000US-0246474P.
PR 08-NOV-2000; 2000US-0246475P.
PR 08-NOV-2000; 2000US-0246476P.
PR 08-NOV-2000; 2000US-0246477P.
PR 08-NOV-2000; 2000US-0246478P.
PR 08-NOV-2000; 2000US-0246523P.
PR 08-NOV-2000; 2000US-0246524P.
PR 08-NOV-2000; 2000US-0246525P.
PR 08-NOV-2000; 2000US-0246526P.
PR 08-NOV-2000; 2000US-0246527P.
PR 08-NOV-2000; 2000US-0246528P.
PR 08-NOV-2000; 2000US-0246532P.
PR 08-NOV-2000; 2000US-0246609P.
PR 08-NOV-2000; 2000US-0246610P.
PR 08-NOV-2000; 2000US-0246611P.
PR 08-NOV-2000; 2000US-0246613P.
PR 17-NOV-2000; 2000US-0249207P.
PR 17-NOV-2000; 2000US-0249208P.
PR 17-NOV-2000; 2000US-0249209P.
PR 17-NOV-2000; 2000US-0249210P.
PR 17-NOV-2000; 2000US-0249211P.
PR 17-NOV-2000; 2000US-0249212P.
PR 17-NOV-2000; 2000US-0249213P.
PR 17-NOV-2000; 2000US-0249214P.
PR 17-NOV-2000; 2000US-0249215P.
PR 17-NOV-2000; 2000US-0249216P.
PR 17-NOV-2000; 2000US-0249217P.
PR 17-NOV-2000; 2000US-0249218P.
PR 17-NOV-2000; 2000US-0249218P.
PR 17-NOV-2000; 2000US-0249244P.
PR 17-NOV-2000; 2000US-0249245P.
PR 17-NOV-2000; 2000US-0249264P.
PR 17-NOV-2000; 2000US-0249265P.
PR 17-NOV-2000; 2000US-0249297P.
PR 17-NOV-2000; 2000US-0249299P.

PR 17-NOV-2000; 2000US-0249300P.
PR 01-DEC-2000; 2000US-0250160P.
PR 01-DEC-2000; 2000US-0250391P.
PR 05-DEC-2000; 2000US-0251030P.
PR 05-DEC-2000; 2000US-0251988P.
PR 05-DEC-2000; 2000US-0256719P.
PR 06-DEC-2000; 2000US-0251479P.
PR 08-DEC-2000; 2000US-0251856P.
PR 08-DEC-2000; 2000US-0251868P.
PR 08-DEC-2000; 2000US-0251869P.
PR 08-DEC-2000; 2000US-0251989P.
PR 11-DEC-2000; 2000US-0251990P.
PR 05-JAN-2001; 2001US-0259678P.
XX (HUMA-) HUMAN GENOME SCI INC.
XX Rosen CA, Barash SC, Ruben SM;
XX WPI; 2001-502630/55.
XX Polynucleotides encoding digestive system antigens, useful for
PT diagnosing, treating, preventing and/or prognosing disorders of the
PT digestive system, particularly cancer and cancer metastases.
XX Disclosure; SEQ ID NO 3071; 986pp; English.
XX The present invention provides the protein and coding sequences of a
CC number of human digestive system antigens. These can be used in the
CC diagnosis, treatment and prevention of digestive system disorders,
CC including cancer, Meckel's diverticulum, bacterial or parasitic
CC infections, appendicitis, Hirschsprung's disease, chronic colitis or
CC ulcerative colitis. The present sequence is a genomic DNA fragment
CC encoding a digestive system antigen of the invention
XX
SQ Sequence 8417 BP; 2708 A; 1429 C; 1657 G; 2621 T; 0 U; 2 Other;

Alignment Scores:
Pred. No.: 102 Length: 8417
Score: 56.00 Matches: 9
Percent Similarity: 78.57% Conservative: 2
Best Local Similarity: 64.29% Mismatches: 3
Query Match: 56.57% Indels: 0
DB: 4 Gaps: 0

US-10-799-747-116 (1-20) x AAK89495 (1-8417)

Qy 1 MetAlaAlaHisSerValLeuSerPheLeuLeuTrpThrPro 14
Db 7484 ATGCCACACATTCGATTTGACGTTTATTCTGACCCCA 7443

RESULT 5
AAS32681
ID AAS32681 standard; DNA; 8417 BP.
XX
AC AAS32681;
XX
DT 17-DEC-2001 (first entry)
XX
DE Human genomic DNA for novel endocrine antigen, SEQ ID No 635.
XX
KW Human; endocrine antigen; ds; cytostatic; antiinfertility; antidiabetic;
KW thyroid-active; adrenal-active; androgenic; gastric; gene therapy;
KW antitense-therapy; antibody; endocrine disorder; hormone imbalance;
KW reproductive disorder; endocrine cancer; pancreatic disorder;
KW diabetes mellitus; adrenal gland disorder; hirsutism; thyroid disorder;
KW hyperthyroidism; hypothalamic disorder; vanishing testes syndrome.
XX
OS Homo sapiens.
XX
PN WO200155319-A2.
XX
PD 02-AUG-2001.

PR	17-MAR-2000;	2000US-01900761
PR	18-APR-2000;	2000US-01981233
PR	19-MAY-2000;	2000US-0205153P
PR	26-JUN-2000;	2000US-0209467P
PR	28-JUN-2000;	2000US-0214886P
PR	30-JUN-2000;	2000US-0215135P
PR	07-JUL-2000;	2000US-0216647P
PR	07-JUL-2000;	2000US-0216800P
PR	11-JUL-2000;	2000US-0217487P
PR	11-JUL-2000;	2000US-0217496P
PR	14-JUL-2000;	2000US-0218290P
PR	16-JUL-2000;	2000US-0220963P
PR	26-JUL-2000;	2000US-0220964P
PR	14-AUG-2000;	2000US-0224518P
PR	14-AUG-2000;	2000US-0224519P
PR	14-AUG-2000;	2000US-0225213P
PR	14-AUG-2000;	2000US-0225214P
PR	14-AUG-2000;	2000US-0225266P
PR	14-AUG-2000;	2000US-0225267P
PR	14-AUG-2000;	2000US-0225268P
PR	14-AUG-2000;	2000US-0225270P
PR	14-AUG-2000;	2000US-0225447P
PR	14-AUG-2000;	2000US-0225575P
PR	14-AUG-2000;	2000US-0225758P
PR	30-AUG-2000;	2000US-0227009P
PR	31-AUG-2000;	2000US-0228924P
PR	18-AUG-2000;	2000US-0225759P
PR	18-AUG-2000;	2000US-0226279P
PR	22-AUG-2000;	2000US-0226681P
PR	22-AUG-2000;	2000US-0226868P
PR	22-AUG-2000;	2000US-0227182P
PR	23-AUG-2000;	2000US-0227009P
PR	30-AUG-2000;	2000US-0228924P
PR	01-SEP-2000;	2000US-0229287P
PR	01-SEP-2000;	2000US-0229343P
PR	01-SEP-2000;	2000US-0229344P
PR	01-SEP-2000;	2000US-0229345P
PR	03-SEP-2000;	2000US-0229509P
PR	05-SEP-2000;	2000US-0229513P
PR	06-SEP-2000;	2000US-0230437P
PR	06-SEP-2000;	2000US-0230438P
PR	08-SEP-2000;	2000US-0231242P
PR	08-SEP-2000;	2000US-0231243P
PR	08-SEP-2000;	2000US-0231344P
PR	08-SEP-2000;	2000US-0231413P
PR	08-SEP-2000;	2000US-0231414P
PR	08-SEP-2000;	2000US-0232080P
PR	08-SEP-2000;	2000US-0232081P
PR	12-SEP-2000;	2000US-0231968P
PR	14-SEP-2000;	2000US-0232397P
PR	14-SEP-2000;	2000US-0232398P
PR	14-SEP-2000;	2000US-0232399P
PR	14-SEP-2000;	2000US-0232400P
PR	14-SEP-2000;	2000US-0232401P
PR	14-SEP-2000;	2000US-0233063P
PR	14-SEP-2000;	2000US-0233064P
PR	14-SEP-2000;	2000US-0233065P
PR	21-SEP-2000;	2000US-0234223P
PR	21-SEP-2000;	2000US-0234224P
PR	25-SEP-2000;	2000US-0234997P
PR	25-SEP-2000;	2000US-0234998P
PR	26-SEP-2000;	2000US-0235484P
PR	27-SEP-2000;	2000US-0235834P
PR	27-SEP-2000;	2000US-0235836P
PR	29-SEP-2000;	2000US-0236327P
PR	29-SEP-2000;	2000US-0236327P
PR	29-SEP-2000;	2000US-0236368P
PR	29-SEP-2000;	2000US-0236369P
PR	02-OCT-2000;	2000US-0236370P
PR	02-OCT-2000;	2000US-0236802P
PR	02-OCT-2000;	2000US-0237037P
PR	02-OCT-2000;	2000US-0237038P
PR	02-OCT-2000;	2000US-0237039P
PR	13-OCT-2000;	2000US-0237040P
PR	13-OCT-2000;	2000US-0239393P

PR	27-OCT-2000	2000US-02270099
PR	23-AUG-2000	2000US-02270099
PR	30-AUG-2000	2000US-0228924P
PR	01-SEP-2000	2000US-0228924P
PR	01-SEP-2000	2000US-0229287P
PR	01-SEP-2000	2000US-0229343P
PR	01-SEP-2000	2000US-0229344P
PR	01-SEP-2000	2000US-0229345P
PR	05-SEP-2000	2000US-0229509P
PR	05-SEP-2000	2000US-0229513P
PR	06-SEP-2000	2000US-0230437P
PR	06-SEP-2000	2000US-0230437P
PR	08-SEP-2000	2000US-0230438P
PR	08-SEP-2000	2000US-0231242Z
PR	08-SEP-2000	2000US-0231243P
PR	08-SEP-2000	2000US-0231244P
PR	08-SEP-2000	2000US-0231413P
PR	08-SEP-2000	2000US-0231414P
PR	08-SEP-2000	2000US-0232080P
PR	08-SEP-2000	2000US-0232081P
PR	12-SEP-2000	2000US-0231968P
PR	14-SEP-2000	2000US-0232397P
PR	14-SEP-2000	2000US-0232398P
PR	14-SEP-2000	2000US-0232399P
PR	14-SEP-2000	2000US-0232400P
PR	14-SEP-2000	2000US-0232401P
PR	14-SEP-2000	2000US-0233063P
PR	14-SEP-2000	2000US-0233064P
PR	14-SEP-2000	2000US-0233065P
PR	21-SEP-2000	2000US-0234223P
PR	21-SEP-2000	2000US-0234274P
PR	25-SEP-2000	2000US-0234597P
PR	25-SEP-2000	2000US-0234598P
PR	26-SEP-2000	2000US-0235484P
PR	27-SEP-2000	2000US-0235834P
PR	27-SEP-2000	2000US-0235836P
PR	29-SEP-2000	2000US-0236327P
PR	29-SEP-2000	2000US-0236367P
PR	29-SEP-2000	2000US-0236368P
PR	29-SEP-2000	2000US-0236369P
PR	02-OCT-2000	2000US-0236370P
PR	02-OCT-2000	2000US-0236802P
PR	02-OCT-2000	2000US-0237037P
PR	02-OCT-2000	2000US-0237038P
PR	02-OCT-2000	2000US-0237039P
PR	03-OCT-2000	2000US-0237040P
PR	13-OCT-2000	2000US-0239395P

PR	08-SEP-2000	2000US-02314144	PR
PR	08-SEP-2000	2000US-0232080P	PR
PR	08-SEP-2000	2000US-0232081P	PR
PR	12-SEP-2000	2000US-0231968P	PR
PR	14-SEP-2000	2000US-0232397P	PR
PR	14-SEP-2000	2000US-0232398P	PR
PR	14-SEP-2000	2000US-0232399P	PR
PR	14-SEP-2000	2000US-0232400P	PR
PR	14-SEP-2000	2000US-0232401P	PR
PR	14-SEP-2000	2000US-0233063P	PR
PR	14-SEP-2000	2000US-0233064P	PR
PR	14-SEP-2000	2000US-0233065P	PR
PR	21-SEP-2000	2000US-0234223P	PR
PR	21-SEP-2000	2000US-0234274P	PR
PR	25-SEP-2000	2000US-0234597P	PR
PR	25-SEP-2000	2000US-0234998P	PR
PR	27-SEP-2000	2000US-0235484P	PR
PR	27-SEP-2000	2000US-0235834P	PR
PR	29-SEP-2000	2000US-0236336P	PR
PR	29-SEP-2000	2000US-0236327P	PR
PR	29-SEP-2000	2000US-0236367P	PR
PR	29-SEP-2000	2000US-0236368P	PR
PR	29-SEP-2000	2000US-0236369P	PR
PR	29-SEP-2000	2000US-0236370P	PR
PR	02-OCT-2000	2000US-0236802P	PR
PR	02-OCT-2000	2000US-0237037P	PR
PR	02-OCT-2000	2000US-0237038P	PR
PR	02-OCT-2000	2000US-0237039P	PR
PR	02-OCT-2000	2000US-0237040P	PR
PR	13-OCT-2000	2000US-0239935P	PR

PR	14-SEP-2000	2000US-02330646	PR
PR	15-SEP-2000	2000US-02330655	PR
PR	21-SEP-2000	2000US-02345233	PR
PR	21-SEP-2000	2000US-02345237	PR
PR	21-SEP-2000	2000US-02349776	PR
PR	25-SEP-2000	2000US-02349888	PR
PR	26-SEP-2000	2000US-02354848	PR
PR	27-SEP-2000	2000US-02358343	PR
PR	27-SEP-2000	2000US-02358366	PR
PR	29-SEP-2000	2000US-02363276	PR
PR	29-SEP-2000	2000US-02363277	PR
PR	29-SEP-2000	2000US-02363676	PR
PR	29-SEP-2000	2000US-02363688	PR
PR	29-SEP-2000	2000US-02363699	PR
PR	29-SEP-2000	2000US-02363700	PR
PR	02-OCT-2000	2000US-02368026	PR
PR	02-OCT-2000	2000US-02370376	PR
PR	02-OCT-2000	2000US-02370388	PR
PR	02-OCT-2000	2000US-02370399	PR
PR	02-OCT-2000	2000US-02370400	PR
PR	13-OCT-2000	2000US-02379435	PR

PR	27-SEP-2000	2000US-02358344P
PR	27-SEP-2000	2000US-02358346P
PR	27-SEP-2000	2000US-02358347P
PR	28-SEP-2000	2000US-0236327P
PR	28-SEP-2000	2000US-02363276P
PR	29-SEP-2000	2000US-0236367P
PR	29-SEP-2000	2000US-0236368P
PR	28-SEP-2000	2000US-0236369P
PR	28-SEP-2000	2000US-0236370P
PR	02-OCT-2000	2000US-0236802P
PR	02-OCT-2000	2000US-0237037P
PR	02-OCT-2000	2000US-0237038P
PR	02-OCT-2000	2000US-0237039P
PR	02-OCT-2000	2000US-0237040P
PR	13-OCT-2000	2000US-0239935P

PR 13-OCT-2000; 2000US-0239937P.
PR 20-OCT-2000; 2000US-0240960P.
PR 20-OCT-2000; 2000US-0241221P.
PR 20-OCT-2000; 2000US-0241785P.
PR 20-OCT-2000; 2000US-0241786P.
PR 20-OCT-2000; 2000US-0241787P.
PR 20-OCT-2000; 2000US-0241808P.
PR 20-OCT-2000; 2000US-0241809P.
PR 20-OCT-2000; 2000US-0241826P.
PR 01-NOV-2000; 2000US-0244617P.
PR 08-NOV-2000; 2000US-0246474P.
PR 08-NOV-2000; 2000US-0246475P.
PR 08-NOV-2000; 2000US-0246476P.
PR 08-NOV-2000; 2000US-0246477P.
PR 08-NOV-2000; 2000US-0246478P.
PR 08-NOV-2000; 2000US-0246523P.
PR 08-NOV-2000; 2000US-0246524P.
PR 08-NOV-2000; 2000US-0246525P.
PR 08-NOV-2000; 2000US-0248526P.
PR 08-NOV-2000; 2000US-0248527P.
PR 08-NOV-2000; 2000US-0248528P.
PR 08-NOV-2000; 2000US-0248532P.
PR 08-NOV-2000; 2000US-0248609P.
PR 08-NOV-2000; 2000US-0248610P.
PR 08-NOV-2000; 2000US-0248611P.
PR 08-NOV-2000; 2000US-0248613P.
PR 17-NOV-2000; 2000US-0249207P.
PR 17-NOV-2000; 2000US-0249208P.
PR 17-NOV-2000; 2000US-0249209P.
PR 17-NOV-2000; 2000US-0249210P.
PR 17-NOV-2000; 2000US-0249211P.
PR 17-NOV-2000; 2000US-0249212P.
PR 17-NOV-2000; 2000US-0249213P.
PR 17-NOV-2000; 2000US-0249214P.
PR 17-NOV-2000; 2000US-0249215P.
PR 17-NOV-2000; 2000US-0249216P.
PR 17-NOV-2000; 2000US-0249217P.
PR 17-NOV-2000; 2000US-0249218P.
PR 17-NOV-2000; 2000US-0249244P.
PR 17-NOV-2000; 2000US-0249245P.
PR 17-NOV-2000; 2000US-0249264P.
PR 17-NOV-2000; 2000US-0249265P.
PR 17-NOV-2000; 2000US-0249267P.
PR 17-NOV-2000; 2000US-0249299P.
PR 17-NOV-2000; 2000US-0249300P.
PR 01-DEC-2000; 2000US-0250160P.
PR 01-DEC-2000; 2000US-0250391P.
PR 05-DEC-2000; 2000US-0251030P.
PR 05-DEC-2000; 2000US-0251988P.
PR 05-DEC-2000; 2000US-0256719P.
PR 06-DEC-2000; 2000US-0251479P.
PR 06-DEC-2000; 2000US-0251856P.
PR 08-DEC-2000; 2000US-0251868P.
PR 08-DEC-2000; 2000US-0251869P.
PR 08-DEC-2000; 2000US-0251989P.
PR 08-DEC-2000; 2000US-0251990P.
PR 11-DEC-2000; 2000US-0254097P.
PR 05-JAN-2001; 2001US-0259678P.
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX Rosen CA, Barash SC, Ruben SM;
XX WPI; 2001-457726/49.
XX
XX Isolated polypeptide for treating, preventing and prognosing disorders
XX related to the endocrine system including endocrine disorders,
XX reproductive disorders, and gastrointestinal disorders and also for
XX testing and detection e.g. diagnosis.
XX
XX Disclosure; SEQ ID NO 426; 558pp; English.
XX
XX The invention relates to cDNAs encoding novel human endocrine antigens or

CC a fragment having biological activity, a domain, an epitope, full length
CC protein, variant, allelic variant or a species homologue of the
CC cDNA/antigen. The DNAs and polypeptides are useful for preventing,
CC treating or ameliorating a medical condition when administered (e.g. by
CC gene therapy or antisense-therapy). Identifying mutations in the genes
CC coding for the antigens is useful for diagnosing a pathological condition
CC or a susceptibility to a pathological condition. The DNAs, antigens and
CC antibodies raised against the antigens useful for treating, preventing
CC and/or prognosing disorders related to the endocrine system or hormone
CC imbalance or reproductive disorders, cancers of endocrine tissues,
CC disorders of the pancreas (e.g. diabetes mellitus), the adrenal glands
CC (e.g. hirsutism), ovaries, the thyroid (e.g. hyperthyroidism), the
CC hypothalamus and testes (e.g. vanishing testes syndrome), many examples
CC of diseases and disorders are given in the specification. The present
CC sequence is genomic DNA fragment form a gene encoding an endocrine
CC antigen of the invention. Note: The sequence data for this patent did not
CC form part of the printed specification, but was obtained in electronic
CC format directly from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 5760 BP; 1725 A; 1199 C; 1068 G; 1768 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.: 208 Length: 5760
Score: 53.00 Matches: 10
Percent Similarity: 68.75% Conservative: 1
Best Local Similarity: 62.50% Mismatches: 5
Query Match: 53.54% Indels: 0
DB: 4 Gaps: 0

US-10-799-747-116 (1-20) x AAS32472 (1-5760)

QY 2 AlaAlaHisSerValLeuSerPheLeuLeuTrpThrProTyrAlaLeu 17
Db 567 GCCCGGCACCTCAACCTGAGCTTTGGCTTGGGTGTATATCTT 614

RESULT 7
AAS32471
ID AAS32471 standard; DNA; 5763 BP.
XX
AC AAS32471;
XX
DT 17-DEC-2001 (first entry)
XX
DE Human genomic DNA for novel endocrine antigen, SEQ ID No 425.
XX
KW Human; endocrine antigen; ds; cytostatic; antiinfertility; antidiabetic;
KW thyroid-active; adrenal-active; androgenic; gastric; gene therapy;
KW antisense-therapy; antibody; endocrine disorder; hormone imbalance;
KW reproductive disorder; endocrine cancer; pancreatic disorder;
KW diabetes mellitus; adrenal gland disorder; hirsutism; thyroid disorder;
KW hyperthyroidism; hypothalamic disorder; vanishing testes syndrome.
XX
OS Homo sapiens.
XX
PN WO200155319-A2.
XX
PD 02-AUG-2001.
XX
PF 17-JAN-2001; 2001WO-US001335.
XX
PR 31-JAN-2000; 2000US-0179065P.
PR 04-FEB-2000; 2000US-0180628P.
PR 24-FEB-2000; 2000US-0184664P.
PR 02-MAR-2000; 2000US-0186350P.
PR 16-MAR-2000; 2000US-0189874P.
PR 17-MAR-2000; 2000US-0190076P.
PR 18-APR-2000; 2000US-0198123P.
PR 19-MAY-2000; 2000US-0205515P.
PR 07-JUN-2000; 2000US-0209467P.
PR 28-JUN-2000; 2000US-0214886P.
PR 30-JUN-2000; 2000US-0215135P.
PR 07-JUL-2000; 2000US-0216647P.
PR 07-JUL-2000; 2000US-0216880P.

PR 11-JUL-2000; 2000US-02117487P.
PR 11-JUL-2000; 2000US-0217496P.
PR 14-JUL-2000; 2000US-0218293P.
PR 26-JUL-2000; 2000US-0220963P.
PR 26-JUL-2000; 2000US-0220964P.
PR 14-AUG-2000; 2000US-0224518P.
PR 14-AUG-2000; 2000US-0224519P.
PR 14-AUG-2000; 2000US-0225213P.
PR 14-AUG-2000; 2000US-0225214P.
PR 14-AUG-2000; 2000US-0225266P.
PR 14-AUG-2000; 2000US-0225267P.
PR 14-AUG-2000; 2000US-0225268P.
PR 14-AUG-2000; 2000US-0225270P.
PR 14-AUG-2000; 2000US-0225271P.
PR 14-AUG-2000; 2000US-0225447P.
PR 14-AUG-2000; 2000US-0225757P.
PR 14-AUG-2000; 2000US-0225758P.
PR 14-AUG-2000; 2000US-0225759P.
PR 18-AUG-2000; 2000US-0226279P.
PR 22-AUG-2000; 2000US-0226681P.
PR 22-AUG-2000; 2000US-0226868P.
PR 22-AUG-2000; 2000US-0227182P.
PR 23-AUG-2000; 2000US-0227009P.
PR 30-AUG-2000; 2000US-0228924P.
PR 01-SEP-2000; 2000US-0229287P.
PR 01-SEP-2000; 2000US-0229343P.
PR 01-SEP-2000; 2000US-0229344P.
PR 01-SEP-2000; 2000US-0229345P.
PR 05-SEP-2000; 2000US-0229509P.
PR 05-SEP-2000; 2000US-0229513P.
PR 06-SEP-2000; 2000US-0230437P.
PR 06-SEP-2000; 2000US-0230438P.
PR 08-SEP-2000; 2000US-0231242P.
PR 08-SEP-2000; 2000US-0231243P.
PR 08-SEP-2000; 2000US-0231244P.
PR 08-SEP-2000; 2000US-0231413P.
PR 08-SEP-2000; 2000US-0231414P.
PR 08-SEP-2000; 2000US-0232080P.
PR 08-SEP-2000; 2000US-0232081P.
PR 12-SEP-2000; 2000US-0231968P.
PR 14-SEP-2000; 2000US-0232397P.
PR 14-SEP-2000; 2000US-0232398P.
PR 14-SEP-2000; 2000US-0232399P.
PR 14-SEP-2000; 2000US-0232400P.
PR 14-SEP-2000; 2000US-0232401P.
PR 14-SEP-2000; 2000US-0233053P.
PR 14-SEP-2000; 2000US-0233064P.
PR 14-SEP-2000; 2000US-0233065P.
PR 21-SEP-2000; 2000US-0234223P.
PR 21-SEP-2000; 2000US-0234274P.
PR 25-SEP-2000; 2000US-0234997P.
PR 25-SEP-2000; 2000US-0234998P.
PR 26-SEP-2000; 2000US-0235484P.
PR 27-SEP-2000; 2000US-0235834P.
PR 27-SEP-2000; 2000US-0235836P.
PR 29-SEP-2000; 2000US-0236327P.
PR 29-SEP-2000; 2000US-0236327P.
PR 29-SEP-2000; 2000US-0236367P.
PR 29-SEP-2000; 2000US-0236368P.
PR 29-SEP-2000; 2000US-0236369P.
PR 29-SEP-2000; 2000US-0236370P.
PR 02-OCT-2000; 2000US-0236802P.
PR 02-OCT-2000; 2000US-0237037P.
PR 02-OCT-2000; 2000US-0237038P.
PR 02-OCT-2000; 2000US-0237039P.
PR 02-OCT-2000; 2000US-0237040P.
PR 13-OCT-2000; 2000US-0239935P.
PR 13-OCT-2000; 2000US-0239937P.
PR 20-OCT-2000; 2000US-0240960P.
PR 20-OCT-2000; 2000US-0241221P.
PR 20-OCT-2000; 2000US-0241785P.
PR 20-OCT-2000; 2000US-0241786P.
PR 20-OCT-2000; 2000US-0241787P.
PR 20-OCT-2000; 2000US-0241808P.
PR 20-OCT-2000; 2000US-0241809P.

PR 20-OCT-2000; 2000US-0241826P.
PR 01-NOV-2000; 2000US-0244617P.
PR 08-NOV-2000; 2000US-0246474P.
PR 08-NOV-2000; 2000US-0246475P.
PR 08-NOV-2000; 2000US-0246476P.
PR 08-NOV-2000; 2000US-0246477P.
PR 08-NOV-2000; 2000US-0246478P.
PR 08-NOV-2000; 2000US-0246523P.
PR 08-NOV-2000; 2000US-0246524P.
PR 08-NOV-2000; 2000US-0246525P.
PR 08-NOV-2000; 2000US-0246526P.
PR 08-NOV-2000; 2000US-0246527P.
PR 08-NOV-2000; 2000US-0246528P.
PR 08-NOV-2000; 2000US-0246532P.
PR 08-NOV-2000; 2000US-0246609P.
PR 08-NOV-2000; 2000US-0246610P.
PR 08-NOV-2000; 2000US-0246611P.
PR 08-NOV-2000; 2000US-0246613P.
PR 17-NOV-2000; 2000US-0249207P.
PR 17-NOV-2000; 2000US-0249208P.
PR 17-NOV-2000; 2000US-0249209P.
PR 17-NOV-2000; 2000US-0249210P.
PR 17-NOV-2000; 2000US-0249211P.
PR 17-NOV-2000; 2000US-0249212P.
PR 17-NOV-2000; 2000US-0249213P.
PR 17-NOV-2000; 2000US-0249214P.
PR 17-NOV-2000; 2000US-0249215P.
PR 17-NOV-2000; 2000US-0249216P.
PR 17-NOV-2000; 2000US-0249217P.
PR 17-NOV-2000; 2000US-0249218P.
PR 17-NOV-2000; 2000US-0249244P.
PR 17-NOV-2000; 2000US-0249245P.
PR 17-NOV-2000; 2000US-0249264P.
PR 17-NOV-2000; 2000US-0249265P.
PR 17-NOV-2000; 2000US-0249297P.
PR 17-NOV-2000; 2000US-0249299P.
PR 17-NOV-2000; 2000US-0249300P.
PR 01-DEC-2000; 2000US-0250160P.
PR 01-DEC-2000; 2000US-0250391P.
PR 05-DEC-2000; 2000US-0251030P.
PR 05-DEC-2000; 2000US-0251988P.
PR 05-DEC-2000; 2000US-0256719P.
PR 06-DEC-2000; 2000US-0251479P.
PR 08-DEC-2000; 2000US-0251856P.
PR 08-DEC-2000; 2000US-0251868P.
PR 08-DEC-2000; 2000US-0251869P.
PR 08-DEC-2000; 2000US-0251989P.
PR 08-DEC-2000; 2000US-0251990P.
PR 11-DEC-2000; 2000US-0254097P.
PR 05-JAN-2001; 2001US-0259678P.
XX
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX Rosen CA, Barash SC, Ruben SM;
XX WPI; 2001-457726/49.
XX
XX Isolated polypeptide for treating, preventing and prognosing disorders
XX related to the endocrine system including endocrine disorders,
XX reproductive disorders, and gastrointestinal disorders and also for
XX testing and detection e.g. diagnosis.
XX
XX Disclosure; SEQ ID NO 425; 558pp; English.
XX
XX The invention relates to cDNAs encoding novel human endocrine antigens or
XX a fragment having biological activity, a domain, an epitope, full length
XX protein, variant, allelic variant or a species homologue of the
XX cDNA/antigen. The DNAs and polypeptides are useful for preventing,
XX treating or ameliorating a medical condition when administered (e.g. by
XX gene therapy or antisense-therapy). Identifying mutations in the genes
XX coding for the antigens is useful for diagnosing a pathological condition
XX or a susceptibility to a pathological condition. The DNAs, antigens and
XX antibodies raised against the antigens useful for treating, preventing

CC and/ or prognosing disorders related to the endocrine system or hormone
 CC imbalance or reproductive disorders, cancers of endocrine tissues,
 CC disorders of the pancreas (e.g. diabetes mellitus), the adrenal glands
 CC (e.g. hirsutism), ovaries, the thyroid (e.g. hyperthyroidism), the
 CC hypothalamus and testes (e.g. vanishing testes syndrome), many examples
 CC of diseases and disorders are given in the specification. The present
 CC sequence is genomic DNA fragment form a gene encoding an endocrine
 CC antigen of the invention. Note: The sequence data for this patent did not
 CC form part of the printed specification, but was obtained in electronic
 CC format directly from WIPO at ftp.wipo.int/pub/published_pct_sequences
 XX
 SQ Sequence 5763 BP; 1725 A; 1203 C; 1066 G; 1768 T; 0 U; 1 Other;

Alignment Scores:
 Pred. No.: 209 Length: 5763
 Score: 53.00 Matches: 10
 Percent Similarity: 68.75% Conservative: 1
 Best Local Similarity: 62.50% Mismatches: 5
 Query Match: 53.54% Indels: 0
 DB: 4 Gaps: 0
 US-10-799-747-116 (1-20) x AAS32471 (1-5763)
 QY 2 AlaAlaHisSerValLeuSerPheLeuLeuTTPThrProTyrAlaLeu 17
 Db 567 GCCCGCAGCTCAACCTGAGCTTTCCCTTGGTGTATATCTCTT 614

RESULT 8
 ABZ36220
 ID ABZ36220 standard; cDNA; 2877 BP.
 AC ABZ36220;
 XX
 DT 10-FEB-2003 (first entry)
 XX
 DE Human secretory polynucleotide SPTM SEQ ID NO 384.
 XX

KW Human; SPTM; autoimmune disorder; inflammatory disorder; AIDS; anaemia;
 KW asthma; Crohn's disease; neurological disorder; epilepsy; cancer;
 KW Huntington's disease; Alzheimer's disease; Creutzfeldt-Jakob disease;
 KW multiple sclerosis; Parkinson's disease; cell proliferative disorder;
 KW anti-inflammatory; immunosuppressive; neuroprotective; nootropic;
 KW neuroleptic; anticonvulsant; cytostatic; antiparkinsonian; anxiolytic;
 KW antipsoriatic; anticancer; anti-HIV; human immunodeficiency virus;
 KW secretory polynucleotide; secretory protein; gene; ss.

XX Homo sapiens.
 OS
 PN WO200283876-A2.
 XX
 PD 24-OCT-2002.
 XX
 PF 27-MAR-2002; 2002WO-US009921.
 XX
 PR 29-MAR-2001; 2001US-0280067P.
 PR 29-MAR-2001; 2001US-0280068P.
 PR 16-MAY-2001; 2001US-0291280P.
 PR 17-MAY-2001; 2001US-0291829P.
 PR 17-MAY-2001; 2001US-0291849P.
 PR 19-JUN-2001; 2001US-0299428P.
 PR 20-JUN-2001; 2001US-0299776P.
 PR 20-JUN-2001; 2001US-030001P.
 XX
 PA (INCY-) INCYTE GENOMICS INC.

XX Daffo A, Jones AL, Tran AB, Dahl CR, Gietzen D, Chinn J;
 PI Dufour GE, Hillman JL, Yu JY, Thason O, Yap PE, Anshey SR;
 PI Daugherty SC, Dam TC, Liu IF, Nguyen DA, Kleefeld Y, Gerstin EH;
 PI Peralta CH, David ME, Lewis SA, Chen AJ, Panzer SR, Harris B;
 PI Flores V, Marwaha R, Lo A, Lan RY, Urashka ME;
 XX WPI; 2003-075543/07.
 DR P-PSDB; ABP75778.

XX
 PT New human secretory proteins and polynucleotides, useful for diagnosing,
 PT treating or preventing autoimmune/inflammatory disorders (e.g. AIDS),
 PT neurological disorders (e.g. Alzheimer's), or cell proliferations or
 PT cancers.
 XX
 PS Claim 1; SEQ ID NO 384; 458pp + Sequence Listing; English.
 XX
 CC The invention relates to a secretory polynucleotide (designated sptm)
 CC comprising any of 567 polynucleotide sequences (ABZ35837-ABZ36403), a
 CC naturally occurring polynucleotide sequence at least 90 % identical to
 CC the polynucleotide sequence, a polynucleotide complementary to them or an
 CC RNA equivalent of them. The polypeptide or polynucleotide are useful for
 CC treating, preventing or diagnosing a disease or condition associated with
 CC the expression of functional SPTM. These are particularly useful for
 CC diagnosing, treating or preventing autoimmune/inflammatory disorders
 CC (e.g. acquired immunodeficiency syndrome, anaemia, asthma or Crohn's
 CC disease), neurological disorders (e.g. epilepsy, Huntington's disease,
 CC dementia, stroke, Alzheimer's disease, Creutzfeldt-Jakob disease,
 CC multiple sclerosis, cerebral palsy, Parkinson's disease, anxiety,
 CC schizophrenia or amnesia), or cell proliferative disorders (e.g.
 CC psoriasis, polycythemia vera, or cancers including adenocarcinoma,
 CC leukaemia, lymphoma, melanoma, myeloma, sarcoma or cancers of the brain,
 CC breast, cervix or prostate). Note: The sequence data for this patent did
 CC not form part of the printed specification, but was obtained in electronic
 CC format directly from WIPO at ftp.wipo.int/pub/published_pct_sequences
 XX
 SQ Sequence 2877 BP; 854 A; 609 C; 504 G; 910 T; 0 U; 0 Other;

Alignment Scores:
 Pred. No.: 131 Length: 2877
 Score: 52.00 Matches: 9
 Percent Similarity: 73.33% Conservative: 2
 Best Local Similarity: 60.00% Mismatches: 4
 Query Match: 52.53% Indels: 0
 DB: 7 Gaps: 0

US-10-799-747-116 (1-20) x ABZ36220 (1-2877)

QY 2 AlaAlaHisSerValLeuSerPheLeuLeuTTPThrProTyrAla 16
 Db 2193 GCTCTCAGCAAAATATCATTTATTTGGTGCCACAGGCC 2237

RESULT 9
 AAK82208/c
 ID AAK82208 standard; DNA; 36785 BP.
 AC AAK82208;
 XX
 DT 07-NOV-2001 (first entry)
 XX
 DE Human immune/haematopoietic antigen genomic sequence SEQ ID NO:37020.
 KW Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;
 KW cytostatic; gene therapy; vaccine; metastasis; ds.
 XX Homo sapiens.
 OS
 PN WO200157182-A2.
 XX
 PD 09-AUG-2001.
 XX
 PF 17-JAN-2001; 2001WO-US001354.
 XX
 PR 31-JAN-2000; 2000US-0179065P.
 PR 04-FEB-2000; 2000US-0180628P.
 PR 24-FEB-2000; 2000US-0184664P.
 PR 02-MAR-2000; 2000US-0186350P.
 PR 16-MAR-2000; 2000US-0189874P.
 PR 17-MAR-2000; 2000US-0190076P.
 PR 18-APR-2000; 2000US-0198123P.
 PR 19-MAY-2000; 2000US-0205515P.
 PR 07-JUN-2000; 2000US-0209467P.

PR 28-JUN-2000; 2000US-0214886P.
PR 30-JUN-2000; 2000US-0215135P.
PR 07-JUL-2000; 2000US-0216647P.
PR 07-JUL-2000; 2000US-0216880P.
PR 11-JUL-2000; 2000US-0217487P.
PR 11-JUL-2000; 2000US-0217496P.
PR 14-JUL-2000; 2000US-0218290P.
PR 26-JUL-2000; 2000US-0220963P.
PR 26-JUL-2000; 2000US-0220964P.
PR 14-AUG-2000; 2000US-0224518P.
PR 14-AUG-2000; 2000US-0224519P.
PR 14-AUG-2000; 2000US-0225213P.
PR 14-AUG-2000; 2000US-0225214P.
PR 14-AUG-2000; 2000US-0225266P.
PR 14-AUG-2000; 2000US-0225267P.
PR 14-AUG-2000; 2000US-0225268P.
PR 14-AUG-2000; 2000US-0225270P.
PR 14-AUG-2000; 2000US-0225447P.
PR 14-AUG-2000; 2000US-0225757P.
PR 14-AUG-2000; 2000US-0225758P.
PR 14-AUG-2000; 2000US-0225759P.
PR 18-AUG-2000; 2000US-0226279P.
PR 22-AUG-2000; 2000US-0226681P.
PR 22-AUG-2000; 2000US-0226868P.
PR 22-AUG-2000; 2000US-0227182P.
PR 23-AUG-2000; 2000US-0227009P.
PR 30-AUG-2000; 2000US-0228924P.
PR 01-SEP-2000; 2000US-0229287P.
PR 01-SEP-2000; 2000US-0229343P.
PR 01-SEP-2000; 2000US-0229344P.
PR 01-SEP-2000; 2000US-0229345P.
PR 05-SEP-2000; 2000US-0229509P.
PR 05-SEP-2000; 2000US-0229513P.
PR 06-SEP-2000; 2000US-0230437P.
PR 06-SEP-2000; 2000US-0230438P.
PR 08-SEP-2000; 2000US-0231242P.
PR 08-SEP-2000; 2000US-0231243P.
PR 08-SEP-2000; 2000US-0231244P.
PR 08-SEP-2000; 2000US-0231413P.
PR 08-SEP-2000; 2000US-0231414P.
PR 08-SEP-2000; 2000US-0232080P.
PR 08-SEP-2000; 2000US-0232081P.
PR 12-SEP-2000; 2000US-0231968P.
PR 14-SEP-2000; 2000US-0232397P.
PR 14-SEP-2000; 2000US-0232398P.
PR 14-SEP-2000; 2000US-0232399P.
PR 14-SEP-2000; 2000US-0232400P.
PR 14-SEP-2000; 2000US-0232401P.
PR 14-SEP-2000; 2000US-0233063P.
PR 14-SEP-2000; 2000US-0233064P.
PR 14-SEP-2000; 2000US-0233065P.
PR 21-SEP-2000; 2000US-0234223P.
PR 21-SEP-2000; 2000US-0234224P.
PR 25-SEP-2000; 2000US-0234274P.
PR 25-SEP-2000; 2000US-0234997P.
PR 25-SEP-2000; 2000US-0234998P.
PR 26-SEP-2000; 2000US-0235484P.
PR 27-SEP-2000; 2000US-0235834P.
PR 27-SEP-2000; 2000US-0235836P.
PR 29-SEP-2000; 2000US-0236327P.
PR 29-SEP-2000; 2000US-0236367P.
PR 29-SEP-2000; 2000US-0236368P.
PR 29-SEP-2000; 2000US-0236369P.
PR 29-SEP-2000; 2000US-0236370P.
PR 02-OCT-2000; 2000US-0236802P.
PR 02-OCT-2000; 2000US-0237037P.
PR 02-OCT-2000; 2000US-0237038P.
PR 02-OCT-2000; 2000US-0237039P.
PR 02-OCT-2000; 2000US-0237040P.
PR 13-OCT-2000; 2000US-0239935P.
PR 13-OCT-2000; 2000US-0239937P.
PR 20-OCT-2000; 2000US-0240960P.
PR 20-OCT-2000; 2000US-0241221P.
PR 20-OCT-2000; 2000US-0241785P.

PR 20-OCT-2000; 2000US-0241786P.
PR 20-OCT-2000; 2000US-0241787P.
PR 20-OCT-2000; 2000US-0241808P.
PR 20-OCT-2000; 2000US-0241809P.
PR 20-OCT-2000; 2000US-0241826P.
PR 01-NOV-2000; 2000US-0244617P.
PR 08-NOV-2000; 2000US-0246474P.
PR 08-NOV-2000; 2000US-0246475P.
PR 08-NOV-2000; 2000US-0246476P.
PR 08-NOV-2000; 2000US-0246477P.
PR 08-NOV-2000; 2000US-0246478P.
PR 08-NOV-2000; 2000US-0246523P.
PR 08-NOV-2000; 2000US-0246524P.
PR 08-NOV-2000; 2000US-0246525P.
PR 08-NOV-2000; 2000US-0246526P.
PR 08-NOV-2000; 2000US-0246527P.
PR 08-NOV-2000; 2000US-0246528P.
PR 08-NOV-2000; 2000US-0246532P.
PR 08-NOV-2000; 2000US-0246609P.
PR 08-NOV-2000; 2000US-0246610P.
PR 08-NOV-2000; 2000US-0246611P.
PR 08-NOV-2000; 2000US-0246613P.
PR 17-NOV-2000; 2000US-0249207P.
PR 17-NOV-2000; 2000US-0249208P.
PR 17-NOV-2000; 2000US-0249209P.
PR 17-NOV-2000; 2000US-0249210P.
PR 17-NOV-2000; 2000US-0249211P.
PR 17-NOV-2000; 2000US-0249212P.
PR 17-NOV-2000; 2000US-0249213P.
PR 17-NOV-2000; 2000US-0249214P.
PR 17-NOV-2000; 2000US-0249215P.
PR 17-NOV-2000; 2000US-0249216P.
PR 17-NOV-2000; 2000US-0249217P.
PR 17-NOV-2000; 2000US-0249218P.
PR 17-NOV-2000; 2000US-0249244P.
PR 17-NOV-2000; 2000US-0249245P.
PR 17-NOV-2000; 2000US-0249264P.
PR 17-NOV-2000; 2000US-0249265P.
PR 17-NOV-2000; 2000US-0249297P.
PR 17-NOV-2000; 2000US-0249299P.
PR 17-NOV-2000; 2000US-0249300P.
PR 01-DEC-2000; 2000US-0250160P.
PR 01-DEC-2000; 2000US-0250391P.
PR 05-DEC-2000; 2000US-0251030P.
PR 05-DEC-2000; 2000US-0251988P.
PR 05-DEC-2000; 2000US-0256719P.
PR 06-DEC-2000; 2000US-0251479P.
PR 08-DEC-2000; 2000US-0251856P.
PR 08-DEC-2000; 2000US-0251868P.
PR 08-DEC-2000; 2000US-0251869P.
PR 08-DEC-2000; 2000US-0251989P.
PR 08-DEC-2000; 2000US-0251990P.
PR 11-DEC-2000; 2000US-0254097P.
PR 05-JAN-2001; 2001US-0259678P.

(HUMA-) HUMAN GENOME SCI INC.

Rosen CA, Barash SC, Ruben SM;

WPI; 2001-483426/52.

Nucleic acids encoding human immune/hematopoietic antigen polypeptides, useful for preventing, diagnosing and/or treating cancers and metastasis.

Disclosure; SEQ ID NO 37020; 3071pp + Sequence Listing; English.

AAK54951 to AAK64702 encode the human immune/haematopoietic antigen (I) amino acid sequences given in AAM82170 to AAM91921. (I) have cytostatic activity, and can be used in gene therapy and vaccine production. (I) proteins and polynucleotides may be used in the prevention, diagnosis and treatment of diseases associated with inappropriate (I) expression. For example, they may be used to treat disorders associated with decreased expression by rectifying mutations or deletions in a patient's genome

CC that affect the activity of (I) by expressing inactive proteins or to
 CC supplement the patients own production of (I). Additionally, (I)
 CC polynucleotides may be used to produce the secreted (I), by inserting the
 CC nucleic acids into a host cell and culturing the cell to express the
 CC protein. (I) proteins and polynucleotides may be used to prevent,
 CC diagnose and treat immune/haematopoietic-related diseases, especially
 CC cancers and cancer metastases of haematopoietic-derived cells. AAK64703
 CC to AAK87694 represent human immune/haematopoietic antigen genomic
 CC sequences from the present invention. AAK54942 to AAK54950 and AAK82169
 CC represent sequences used in the exemplification of the present invention
 XX

SQ Sequence 36785 BP; 9938 A; 8156 C; 8241 G; 10450 T; 0 U; 0 Other;

Alignment Scores:
 Pred. No.: 3.04e+03 Length: 36785
 Score: 52.00 Matches: 9
 Percent Similarity: 73.33% Conservative: 2
 Best Local Similarity: 60.00% Mismatches: 4
 Query Match: 52.53% Indels: 0
 DB: 4 Gaps: 0

US-10-799-747-116 (1-20) x AAK82208 (1-36785)

QY 2 AlaAlaHisSerValLeuSerPheLeuLeuTyrThrProTyrAla 16

Db 13072 GCTCCTCACAGCAATATCATTCATATTTGGTGCCAGGCC 13028

RESULT 10

AAAF64369/c

ID AAF64369 standard; cDNA; 400 BP.

XX AAF64369;

XX 09-APR-2001 (first entry)

DT 09-APR-2001 (first entry)

XX Novel human polynucleotide, SEQ ID NO: 125.

DE Human; cytostatic; gene therapy; colon cancer; prostate cancer;

XX breast cancer; lung cancer; cancer detection; ss.

XX Homo sapiens.

XX WO200102568-A2.

XX 11-JAN-2001.

PD 30-JUN-2000; 2000WO-US018374.

XX 02-JUL-1999; 99US-0142310P.

XX 02-JUL-1999; 99US-0142311P.

XX (CHIR) CHIRON CORP.

PA (HYSE-) HYSEQ INC.

XX Williams LT, Escobedo J, Inniss MA, Garcia PD, Klinger J;

PI Kassam A, Reinhard C, Randazzo E, Kennedy GC, Pot D, Lamson G;

PI Drmanac R, C-kenjakov R, Drmanac S, Dickson M, Labat I;

PI Leshkowitz D, Kita D, Garcia V, Jones LW, Strache-Crain B;

XX WPI; 2001-091805/10.

DR Library of polynucleotides for diagnosing a cancerous state of a

PT mammalian cell and detecting cancer, particularly of the colon or

PT prostate, comprises 3351 human polynucleotide sequences.

XX Claim 9; Page 562; 1046pp; English.

XX The present sequence is one of 3351 sequences in a library of human

CC polynucleotides. The library is used to detect differentially expressed

CC genes correlated with a cancerous state of a mammalian cell and can

CC detect colon, prostate, breast and lung cancer. The library can be used

CC to produce probes for detection of mRNA and to produce additional copies

CC of the polynucleotides. The probes can be used for chromosome mapping of

CC the polynucleotide and for detection of transcription levels. Ribozymes
 CC or antisense oligonucleotides can be generated. The polynucleotides and
 CC their gene products are used as genetic or biochemical markers (e.g. in
 CC blood or tissues) that will detect the earliest changes along the
 CC carcinogenesis pathway and/or monitor the efficacy of therapies and
 CC preventive interventions. The polynucleotides, polypeptides and
 CC antibodies against them can be used in pharmaceutical compositions to
 CC treat the cancers and proliferative disorders such as neoplasia,
 CC dysplasia and hyperplasia
 XX

SQ Sequence 400 BP; 123 A; 64 C; 119 G; 94 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.: 17.2 Length: 400
 Score: 51.00 Matches: 11
 Percent Similarity: 70.00% Conservative: 3
 Best Local Similarity: 55.00% Mismatches: 4
 Query Match: 51.52% Indels: 2
 DB: 5 Gaps: 1

US-10-799-747-116 (1-20) x AAF64369 (1-400)

QY 2 AlaAlaHisSerValLeuSer-----PheLeuLeuTyrThrProTyrAlaLeuLysSer 19

Db 198 GCTCCCACTCTGTGCAAGTGATCATTCTTCATTTGGACGCCCATTCGACTCTCTGCA 139

RESULT 11

AAH03292/c

ID AAH03292 standard; cDNA; 505 BP.

XX AAH03292;

XX 26-JUN-2001 (first entry)

DT Human cDNA clone (5'-primer) SEQ ID NO:127.

XX Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.

XX Homo sapiens.

XX EP1074617-A2.

XX 07-FEB-2001.

XX 28-JUL-2000; 2000EP-00116126.

XX 29-JUL-1999; 99JP-00248036.

XX 27-AUG-1999; 99JP-00300253.

XX 11-JAN-2000; 2000JP-00118776.

XX 02-MAY-2000; 2000JP-00183767.

XX 09-JUN-2000; 2000JP-00241899.

XX (HELI-) HELIX RES INST.

XX Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;

PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;

XX WPI; 2001-318749/34.

DR Primer sets for synthesizing polynucleotides, particularly the 5602 full-

XX length cDNAs defined in the specification, and for the detection and/or

PT diagnosis of the abnormality of the proteins encoded by the full-length

PT cDNAs.

XX Claim 1; SEQ ID NO 127; 2537pp + Sequence Listing; English.

XX The present invention describes primer sets for synthesizing 5602 full-

CC length cDNAs defined in the specification. Where a primer set comprises:

CC (a) an oligo-dT primer and an oligonucleotide complementary to the

CC complementary strand of a polynucleotide which comprises one of the 5602

CC nucleotide sequences defined in the specification, where the

CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination

CC of an oligonucleotide comprising a sequence complementary to the

polynucleotides. The library is used to detect differentially expressed genes correlated with a cancerous state of a mammalian cell and can detect colon, prostate, breast and lung cancer. The library can be used to produce probes for detection of mRNA and to produce additional copies of the polynucleotides. The probes can be used for chromosome mapping of the polynucleotide and for detection of transcription levels. Ribozymes the polynucleotide and/or monitor the efficacy of therapies and their sense oligonucleotides can be generated. The polynucleotides and/or antisense oligonucleotides are used as genetic or biochemical markers (e.g. in blood or tissues) that will detect the earliest changes along the carcinogenesis pathway and/or monitor the efficacy of therapies and preventive interventions. The polynucleotides, polypeptides and antibodies against them can be used in pharmaceutical compositions to treat the cancers and proliferative disorders such as neoplasia, dysplasia and hyperplasia

Alignment Scores:		
Pred. No.:	34.7	Length: 708
Score:	51.00%	Matches: 11
Percent Similarity:	70.00%	Conservative: 3
Best Local Similarity:	55.00%	Mismatches: 4
Query Match:	51.52%	Indels: 2
DB:	5	Gaps: 1

US-10-799-747-116 (1-20) x AAF64473 (1-708)

Qy 2 AlaAlaHisSerValLeuSer-----PheLeuLeuTrpThrProTyrAlaAlaLeuLysSer 19
nb |||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
202 GGCCTCCGACCTGTGGTGAAGATGATCACTTCCTCATTTTGACCGCCATTGCCACTCTGTGCA 143

RESULT 13

AAT86839
ID AAT86839 standard: cDNA; 1544 BP.

AC AAT86839;

DT 19-DEC-1997 (first entry)

xx
DE cdNA encoding rat CC chemokine receptor.

rat: CC chemokine receptor: screen: binding: ligand; ds.

50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878

EH	KEY	LOCATION/QUANTITY
ET	CNS	33 1100

$$\frac{F}{F_0} = \frac{1}{1 + \frac{1}{2} \frac{F_0}{F_0}} = \frac{1}{1 + \frac{1}{2} \frac{F_0}{F_0}}$$

PN JP09221599-A.

02-SEP-70

PF 22-FEB-1996; 96JP-00035192.

PR 22-FEB-1996; 96JP-00035192.

PA (TAKE) TAKEDA CHEM IND LTD.

DR WPI: 1997-486426/45.

CC chemokine receptor protein - useful to screen for novel binding

PT CC compounds

XX
XX

cc or advertiser

Sequence 1544 BP: 475 A: 347 C: 330 G: 452 T: 0 U: 0 Other:

CC The present sequence is one of 3351 sequences in a library of human

Wed Jul 21 09:11:30 2004

Db 1453 TTGAGCTTTATACCTTTGGTCCCTACTCCTTG 1421

Search completed: July 21, 2004, 01:43:18
Job time : 356 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: July 21, 2004, 01:24:25 ; Search time 72 Seconds

(without alignments)

154.153 Million cell updates/sec

Title:

US-10-799-747-116

Perfect score: 99

Sequence:

1 MAHSVLSFLLTPTVALKXSX 20

Scoring table:

BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched:

682709 seqs, 277475446 residues

Total number of hits satisfying chosen parameters: 1365418

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DEV=xlp
-O=/cgn2_1/USPTO spooll p/US10799747/runat 19072004 161418 21467/app query.fasta_1.199
-DB=Issued Patents NA -QFMT=fastap -SUFFIX=rni -MINMATCH=0.1 -LOOPL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human0.cdi
-LIST=45 -DOCALLIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10799747 @CGN 1.105 @runat 19072004 161418 21467 -NCPU=6 -ICPU=3
-NO WMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPEL=OCK=100 -LONGLOG
-DRV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

Issued Patents NA:*
1: /cgn2_6/prodata/2/ina/5A COMB.seq:*
2: /cgn2_6/prodata/2/ina/5B COMB.seq:*
3: /cgn2_6/prodata/2/ina/6A COMB.seq:*
4: /cgn2_6/prodata/2/ina/6B COMB.seq:*
5: /cgn2_6/prodata/2/ina/PCTUS COMB.seq:*
6: /cgn2_6/prodata/2/ina/backfiles1.seq:*

pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	51	51.5	2500	4	US-09-625-972-20
C 2	51	51.5	9641	4	US-09-625-972-22
C 3	49	49.5	1368	4	US-09-134-000C-1177
C 4	48	48.5	207	4	US-09-313-294A-7069
C 5	48	48.5	1601	3	US-08-492-459-17
C 6	48	48.5	1601	3	US-08-492-459-18
C 7	48	48.5	1601	3	US-08-423-752-17
C 8	48	48.5	1601	3	US-08-423-752-18
C 9	48	48.5	1601	3	US-08-716-873-31
C 10	48	48.5	1601	3	US-08-716-873-32
C 11	48	48.5	1601	3	US-09-368-431-31
C 12	48	48.5	1601	3	US-09-368-431-32

C 13	48	48.5	1601	4	US-09-414-006-17	Sequence 17, Appl
C 14	48	48.5	1601	4	US-09-414-006-18	Sequence 18, Appl
C 15	48	48.5	1601	4	US-09-447-223-17	Sequence 17, Appl
C 16	48	48.5	1601	4	US-09-447-223-18	Sequence 18, Appl
17	48	48.5	2385	3	US-08-492-459-1	Sequence 1, Appl
18	48	48.5	2385	3	US-08-492-459-3	Sequence 3, Appl
19	48	48.5	2385	3	US-08-423-752-1	Sequence 1, Appl
20	48	48.5	2385	3	US-08-423-752-3	Sequence 3, Appl
21	48	48.5	2385	3	US-08-716-873-15	Sequence 15, Appl
22	48	48.5	2385	3	US-08-716-873-17	Sequence 17, Appl
23	48	48.5	2385	3	US-09-368-431-15	Sequence 15, Appl
24	48	48.5	2385	3	US-09-368-431-17	Sequence 17, Appl
25	48	48.5	2385	4	US-09-414-006-1	Sequence 1, Appl
26	48	48.5	2385	4	US-09-414-006-3	Sequence 3, Appl
27	48	48.5	2385	4	US-09-447-223-1	Sequence 1, Appl
28	48	48.5	2385	4	US-09-447-223-3	Sequence 3, Appl
C 29	48	48.5	17000	4	US-09-679-299A-18	Sequence 18, Appl
C 30	47	47.5	2210	4	US-09-016-434-1177	Sequence 1177, Ap
C 31	46	46.5	292	4	US-09-313-294A-4459	Sequence 4459, Ap
C 32	46	46.5	1098	4	US-09-718-692-3	Sequence 3, Appl
C 33	46	46.5	1098	4	US-09-718-852-3	Sequence 3, Appl
C 34	46	46.5	1098	4	US-09-718-815-3	Sequence 3, Appl
C 35	46	46.5	1296	4	US-09-328-352-467	Sequence 467, App
C 36	46	46.5	2595	4	US-09-016-434-1178	Sequence 1178, Ap
C 37	46	46.5	4911	4	US-09-718-692-1	Sequence 1, Appl
C 38	46	46.5	4911	4	US-09-718-852-1	Sequence 1, Appl
C 39	46	46.5	4911	4	US-09-718-815-1	Sequence 1, Appl
C 40	45.5	46.0	1230025	4	US-09-198-452A-1	Sequence 1, Appl
C 41	45	45.5	1495	4	US-09-016-434-1190	Sequence 1190, Ap
C 42	45	45.5	1495	4	US-09-023-655-1021	Sequence 1021, Ap
C 43	45	45.5	1896	4	US-09-134-000C-616	Sequence 616, App
C 44	45	45.5	2156	1	US-08-012-988A-1	Sequence 1, Appl
C 45	45	45.5	2156	4	US-09-023-655-1247	Sequence 1247, Ap

ALIGNMENTS

RESULT 1

US-09-625-972-20/c

; Sequence 20, Application US/09625972

; Patent No. 6566513

; GENERAL INFORMATION:

; APPLICANT: GUERTLER, Lutz Gerhard

; APPLICANT: HAUSER, Hans Peter

; APPLICANT: DONGMO DELOKO, Yvette Beatrice

; APPLICANT: ZEKENG, Leopold

; APPLICANT: KAPTUE, Lazare

; TITLE OF INVENTION: LENTIVIRUS FROM THE GROUP OF IMMUNODEFICIENCY VIRUSES OF DRILL M

; TITLE OF INVENTION: (MANDRILLUS LEUCOPHAUS) AND THEIR USE

; FILE REFERENCE: 067595/0106

; CURRENT APPLICATION NUMBER: US/09/625,972

; CURRENT FILING DATE: 2000-07-29

; PRIOR APPLICATION NUMBER: DE 199 36 003.0

; PRIOR FILING DATE: 1999-08-03

; NUMBER OF SEQ ID NOS: 57

; SOFTWARE: PatentIn version 3.0

; SEQ ID NO 20

; LENGTH: 2500

; TYPE: DNA

; ORGANISM: SIV - viral

US-09-625-972-20

Alignment Scores:

Pred. No.: 36.3 Length: 2500

Score: 51.00 Matches: 8

Percent Similarity: 100.00% Conservative: 3

Best Local Similarity: 72.73% Mismatches: 0

Query Match: 51.52% Indels: 0

DB: 4 Gaps: 0

US-10-799-747-116 (1-20) x US-09-625-972-20 (1-2500)

Qy 7 LeuSerPheLeuLeuTrpThrProTyrAlaLeu 17


```
; TELEPHONE: 202-371-8850
; TELEFAX:
; TELEX:
; INFORMATION FOR SEQ ID NO: 17:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1601
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: genomic DNA
; ANTI-SENSE: Yes
US-08-492-459-17

Alignment Scores:
Pred. No.: 68.2 Length: 1601
Score: 48.00 Matches: 7
Percent Similarity: 64.29% Conservative: 2
Best Local Similarity: 50.00% Mismatches: 5
Query Match: 48.48% Indels: 0
DB: 3 Gaps: 0

US-10-799-747-116 (1-20) x US-08-492-459-17 (1-1601)
QY 4 HisSerValIeuSerPheIeuLeuTIpThrProTyrAlaLeu 17
Db 843 CATAGCATCCTTGATATTITGGCCTGGTTCATATGGAGTC 802

RESULT 6
US-08-492-459-18/c
; Sequence 18, Application US/08492459
; Patent No. 6015689
; GENERAL INFORMATION:
; APPLICANT: Takashi OKADO et al.
; TITLE OF INVENTION: REGULATION OF AUROBASIDIN SENSITIVITY IN FUNGUS
; NUMBER OF SEQUENCES: 36
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Wenderoth, Lind & Ponack
; STREET: 805 Fifteenth Street, N.W., #700
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20005
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.5 inch, 1.4 mb
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: MS-DOS
; SOFTWARE: Wordperfect 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/492,459
; FILING DATE: June 20, 1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/243,403
; FILING DATE: May 16, 1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Warren M. Cheek, Jr.
; REGISTRATION NUMBER: 33,367
; REFERENCE/DOCKET NUMBER:
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-371-8850
; TELEFAX:
; TELEX:
; INFORMATION FOR SEQ ID NO: 18:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1601
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: mRNA
; ANTI-SENSE: Yes
US-08-492-459-18

Alignment Scores:
Pred. No.: 68.2 Length: 1601
Score: 48.00 Matches: 7
Percent Similarity: 64.29% Conservative: 2
Best Local Similarity: 50.00% Mismatches: 5
Query Match: 48.48% Indels: 0
DB: 3 Gaps: 0

US-10-799-747-116 (1-20) x US-08-492-459-18 (1-1601)
QY 4 HisSerValIeuSerPheIeuLeuTIpThrProTyrAlaLeu 17
Db 843 CATAGCATCCTTGATATTITGGCCTGGTTCATATGGAGTC 802

RESULT 7
US-08-423-752-17/c
; Sequence 17, Application US/08423752
; Patent No. 6022949
; GENERAL INFORMATION:
; APPLICANT: Takashi OKADO et al.
; TITLE OF INVENTION: A GENE CODING FOR A PROTEIN REGULATING
; NUMBER OF SEQUENCES: 22
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Wenderoth, Lind & Ponack
; STREET: 805 Fifteenth Street, N.W., #700
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20005
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 5.25 inch, 500 kb
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: MS-DOS
; SOFTWARE: Wordperfect 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/423,752
; FILING DATE: April 18, 1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/243,403
; FILING DATE: May 16, 1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Warren M. Cheek, Jr.
; REGISTRATION NUMBER: 33,367
; REFERENCE/DOCKET NUMBER:
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-371-8850
; TELEFAX:
; TELEX:
; INFORMATION FOR SEQ ID NO: 17:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1601
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: genomic DNA
; ANTI-SENSE: Yes
US-08-423-752-17

Alignment Scores:
Pred. No.: 68.2 Length: 1601
Score: 48.00 Matches: 7
Percent Similarity: 64.29% Conservative: 2
Best Local Similarity: 50.00% Mismatches: 5
Query Match: 48.48% Indels: 0
DB: 3 Gaps: 0

US-10-799-747-116 (1-20) x US-08-423-752-17 (1-1601)
QY 4 HisSerValIeuSerPheIeuLeuTIpThrProTyrAlaLeu 17
Db 843 CATAGCATCCTTGATATTITGGCCTGGTTCATATGGAGTC 802
```

RESULT 8
US-08-423-752-18/c
; Sequence 18, Application US/08423752
; Patent No. 6022949
; GENERAL INFORMATION:
; APPLICANT: Takashi OKADO et al.
; TITLE OF INVENTION: A GENE CODING FOR A PROTEIN REGULATING
; TITLE OF INVENTION: AUROBASIDIN SENSITIVITY
; NUMBER OF SEQUENCES: 22
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Wenderoth, Lind & Ponack
; STREET: 805 Fifteenth Street, N.W., #700
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20005
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 5.25 inch, 500 kb
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: MS-DOS
; SOFTWARE: Wordperfect 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/423,752
; FILING DATE: April 18, 1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/243,403
; FILING DATE: May 16, 1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Warren M. Cheek, Jr.
; REGISTRATION NUMBER: 33,367
; REFERENCE/DOCKET NUMBER:
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-371-8850
; TELEFAX:
; TELEX:
; INFORMATION FOR SEQ ID NO: 18:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1601
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: mRNA
; ANTI-SENSE: Yes
; US-08-423-752-18
Alignment Scores:
Pred. No.: 68.2 Length: 1601
Score: 48.00 Matches: 7
Percent Similarity: 64.29% Conservative: 2
Best Local Similarity: 50.00% Mismatches: 5
Query Match: 48.48% Indels: 0
DB: 3 Gaps: 0
US-10-799-747-116 (1-20) x US-08-423-752-18 (1-1601)
Qy 4 HisSerValLeuSerPheLeuThrProTyrAlaLeu 17
Db 843 CATAGCATCTTGATATTGGCTGGTCCATATGGAGTC 802
RESULT 9
US-08-716-873-31/c
; Sequence 31, Application US/08716873
; Patent No. 6194166
; GENERAL INFORMATION:
; APPLICANT: Takashi OKADO et al.
; TITLE OF INVENTION: GENE REGULATING AUROBASIDIN SENSITIVITY
; NUMBER OF SEQUENCES: 50
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Wenderoth, Lind & Ponack
; STREET: 805 Fifteenth Street, N.W., #700
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20005
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 5.25 inch, 500 kb
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: MS-DOS
; SOFTWARE: Wordperfect 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/716,873
; FILING DATE: September 20, 1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:

; COUNTRY: U.S.A.
; ZIP: 20005
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.5 inch, 1.44 mb
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: MS-DOS
; SOFTWARE: Wordperfect 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/716,873
; FILING DATE: September 20, 1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Warren M. Cheek, Jr.
; REGISTRATION NUMBER: 33,367
; REFERENCE/DOCKET NUMBER:
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-371-8850
; TELEFAX:
; TELEX:
; INFORMATION FOR SEQ ID NO: 31:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1601 bases
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: genomic DNA
; ANTI-SENSE: Yes
; US-08-716-873-31
Alignment Scores:
Pred. No.: 68.2 Length: 1601
Score: 48.00 Matches: 7
Percent Similarity: 64.29% Conservative: 2
Best Local Similarity: 50.00% Mismatches: 5
Query Match: 48.48% Indels: 0
DB: 3 Gaps: 0
US-10-799-747-116 (1-20) x US-08-716-873-31 (1-1601)
Qy 4 HisSerValLeuSerPheLeuThrProTyrAlaLeu 17
Db 843 CATAGCATCTTGATATTGGCTGGTCCATATGGAGTC 802
RESULT 10
US-08-716-873-32/c
; Sequence 32, Application US/08716873
; Patent No. 6194166
; GENERAL INFORMATION:
; APPLICANT: Takashi OKADO et al.
; TITLE OF INVENTION: GENE REGULATING AUROBASIDIN SENSITIVITY
; NUMBER OF SEQUENCES: 50
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Wenderoth, Lind & Ponack
; STREET: 805 Fifteenth Street, N.W., #700
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20005
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.5 inch, 1.44 mb
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: MS-DOS
; SOFTWARE: Wordperfect 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/716,873
; FILING DATE: September 20, 1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:

```

; ATTORNEY/AGENT INFORMATION:
; NAME: Warren M. Cheek, Jr.
; REGISTRATION NUMBER: 33,367
; REFERENCE/DOCKET NUMBER:
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-371-8850
; TELEFAX:
; TELEX:
; INFORMATION FOR SEQ ID NO: 32:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1601 bases
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: mRNA
; ANTI-SENSE: Yes
US-08-716-873-32

Alignment Scores:
Pred. No.: 68.2 Length: 1601
Score: 48.00 Matches: 7
Percent Similarity: 64.29% Conservative: 2
Best Local Similarity: 50.00% Mismatches: 5
Query Match: 48.48% Indels: 0
DB: 3 Gaps: 0

US-10-799-747-116 (1-20) x US-08-716-873-32 (1-1601)

QY 4 HisSerValLeuSerPheLeuLeuThrProTyrAlaLeu 17
Db 843 CATAGCATCCTTGATATTGGCTGGTCCATATGGAGTC 802

RESULT 11
US-09-368-431-31/c
; Sequence 31, Application US/09368431
; Patent No. 6294651
; GENERAL INFORMATION:
; APPLICANT: Takashi OKADO et al.
; TITLE OF INVENTION: GENE REGULATING AUREOBASIDIN SENSITIVITY
; TITLE OF INVENTION: (AS AMENDED)
; NUMBER OF SEQUENCES: 50
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: WENDEROTH, LIND & PONACK, L.L.P.
; STREET: 2033 K Street, N.W., #800
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20006
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.5 inch, 1.44 mb
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: MS-DOS
; SOFTWARE: Wordperfect 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/368,431
; FILING DATE: August 5, 1999
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/716,873
; FILING DATE: September 20, 1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Warren M. Cheek, Jr.
; REGISTRATION NUMBER: 33,367
; REFERENCE/DOCKET NUMBER:
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-721-8200
; TELEFAX: 202-721-8250
; TELEX:
; INFORMATION FOR SEQ ID NO: 31:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1601 bases
; TYPE: nucleic acid
; STRANDEDNESS: double

```

```

; TOPOLOGY: linear
; MOLECULE TYPE: genomic DNA
; ANTI-SENSE: Yes
US-09-368-431-31

Alignment Scores:
Pred. No.: 68.2 Length: 1601
Score: 48.00 Matches: 7
Percent Similarity: 64.29% Conservative: 2
Best Local Similarity: 50.00% Mismatches: 5
Query Match: 48.48% Indels: 0
DB: 3 Gaps: 0

US-10-799-747-116 (1-20) x US-09-368-431-31 (1-1601)

QY 4 HisSerValLeuSerPheLeuLeuThrProTyrAlaLeu 17
Db 843 CATAGCATCCTTGATATTGGCTGGTCCATATGGAGTC 802

RESULT 12
US-09-368-431-32/c
; Sequence 32, Application US/09368431
; Patent No. 6294651
; GENERAL INFORMATION:
; APPLICANT: Takashi OKADO et al.
; TITLE OF INVENTION: GENE REGULATING AUREOBASIDIN SENSITIVITY
; TITLE OF INVENTION: (AS AMENDED)
; NUMBER OF SEQUENCES: 50
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: WENDEROTH, LIND & PONACK, L.L.P.
; STREET: 2033 K Street, N.W., #800
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20006
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.5 inch, 1.44 mb
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: MS-DOS
; SOFTWARE: Wordperfect 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/368,431
; FILING DATE: August 5, 1999
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/716,873
; FILING DATE: September 20, 1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Warren M. Cheek, Jr.
; REGISTRATION NUMBER: 33,367
; REFERENCE/DOCKET NUMBER:
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-721-8200
; TELEFAX: 202-721-8250
; TELEX:
; INFORMATION FOR SEQ ID NO: 32:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1601 bases
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: mRNA
; ANTI-SENSE: Yes
US-09-368-431-32

Alignment Scores:
Pred. No.: 68.2 Length: 1601
Score: 48.00 Matches: 7
Percent Similarity: 64.29% Conservative: 2
Best Local Similarity: 50.00% Mismatches: 5
Query Match: 48.48% Indels: 0
DB: 3 Gaps: 0

```

Wed Jul 21 09:11:30 2004

us-10-799-747-116.rni

US-10-799-747-116 (1-20) x US-09-368-431-32 (1-1601)

Qy 4 HisSerValLeuSerPheLeuLeuTrpThrProTyrAlaLeu 17
Db 843 CATAGCATCCTTGATATTGGCTGGGTTCCATATGGAGTC 802

RESULT 13
US-09-414-006-17/c
; Sequence 17, Application US/09414006
; Patent No. 6348577
; GENERAL INFORMATION:
; APPLICANT: Takashi OKADO et al.
; TITLE OF INVENTION: REGULATION OF AUREOBASIDIN SENSITIVITY (AS
; NUMBER OF SEQUENCES: 36
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Wenderoth, Lind & Ponack, L.L.P.
; STREET: 2033 K Street, N.W., #800
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20006
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.5 inch, 1.4 mb
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: MS-DOS
; SOFTWARE: Wordperfect 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/414,006
; FILING DATE: October 7, 1999
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/492,459
; FILING DATE: June 20, 1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/243,403
; FILING DATE: May 16, 1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Warren M. Cheek, Jr.
; REGISTRATION NUMBER: 33,367
; REFERENCE/DOCKET NUMBER:
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-721-8200
; TELEFAX: 202-721-8250
; TELEX:
; INFORMATION FOR SEQ ID NO: 17:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1601
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: genomic DNA
; ANTI-SENSE: Yes
US-09-414-006-17

Alignment Scores:
Pred. No.: 68.2 Length: 1601
Score: 48.00 Matches: 7
Percent Similarity: 64.29% Conservative: 2
Best Local Similarity: 50.00% Mismatches: 5
Query Match: 48.48% Indels: 0
DB: 4 Gaps: 0
US-09-414-006-17

US-10-799-747-116 (1-20) x US-09-414-006-17 (1-1601)
Qy 4 HisSerValLeuSerPheLeuLeuTrpThrProTyrAlaLeu 17
Db 843 CATAGCATCCTTGATATTGGCTGGGTTCCATATGGAGTC 802

RESULT 14
US-09-414-006-18/c
; Sequence 18, Application US/09414006
; Patent No. 6348577
; GENERAL INFORMATION:
; APPLICANT: Takashi OKADO et al.
; TITLE OF INVENTION: A GENE CODING FOR A PROTEIN REGULATING
; NUMBER OF SEQUENCES: 22
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Wenderoth, Lind & Ponack
; STREET: 805 Fifteenth Street, N.W., #700
; CITY: Washington
; STATE: D.C.

US-10-799-747-116 (1-20) x US-09-414-006-18 (1-1601)
Qy 4 HisSerValLeuSerPheLeuLeuTrpThrProTyrAlaLeu 17
Db 843 CATAGCATCCTTGATATTGGCTGGGTTCCATATGGAGTC 802

RESULT 15
US-09-447-223-17/c
; Sequence 17, Application US/09447223
; Patent No. 6432664
; GENERAL INFORMATION:
; APPLICANT: Takashi OKADO et al.
; TITLE OF INVENTION: A GENE CODING FOR A PROTEIN REGULATING
; NUMBER OF SEQUENCES: 22
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Wenderoth, Lind & Ponack
; STREET: 805 Fifteenth Street, N.W., #700
; CITY: Washington
; STATE: D.C.

US-10-799-747-116 (1-20) x US-09-414-006-18 (1-1601)
Qy 4 HisSerValLeuSerPheLeuLeuTrpThrProTyrAlaLeu 17
Db 843 CATAGCATCCTTGATATTGGCTGGGTTCCATATGGAGTC 802

RESULT 16
US-09-414-006-19/c
; Sequence 19, Application US/09414006
; Patent No. 6348577
; GENERAL INFORMATION:
; APPLICANT: Takashi OKADO et al.
; TITLE OF INVENTION: A GENE CODING FOR A PROTEIN REGULATING
; NUMBER OF SEQUENCES: 22
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Wenderoth, Lind & Ponack
; STREET: 805 Fifteenth Street, N.W., #700
; CITY: Washington
; STATE: D.C.

GENERAL INFORMATION:
; APPLICANT: Takashi OKADO et al.
; TITLE OF INVENTION: REGULATION OF AUREOBASIDIN SENSITIVITY (AS
; NUMBER OF SEQUENCES: 36
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Wenderoth, Lind & Ponack, L.L.P.
; STREET: 2033 K Street, N.W., #800
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20006
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.5 inch, 1.4 mb
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: MS-DOS
; SOFTWARE: Wordperfect 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/414,006
; FILING DATE: October 7, 1999
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/492,459
; FILING DATE: June 20, 1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/243,403
; FILING DATE: May 16, 1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Warren M. Cheek, Jr.
; REGISTRATION NUMBER: 33,367
; REFERENCE/DOCKET NUMBER:
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-721-8200
; TELEFAX: 202-721-8250
; TELEX:
; INFORMATION FOR SEQ ID NO: 18:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1601
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: mrna
; ANTI-SENSE: Yes
US-09-414-006-18

Alignment Scores:
Pred. No.: 68.2 Length: 1601
Score: 48.00 Matches: 7
Percent Similarity: 64.29% Conservative: 2
Best Local Similarity: 50.00% Mismatches: 5
Query Match: 48.48% Indels: 0
DB: 4 Gaps: 0
US-10-799-747-116 (1-20) x US-09-414-006-18 (1-1601)

US-10-799-747-116 (1-20) x US-09-414-006-18 (1-1601)
Qy 4 HisSerValLeuSerPheLeuLeuTrpThrProTyrAlaLeu 17
Db 843 CATAGCATCCTTGATATTGGCTGGGTTCCATATGGAGTC 802

RESULT 16
US-09-447-223-17/c
; Sequence 17, Application US/09447223
; Patent No. 6432664
; GENERAL INFORMATION:
; APPLICANT: Takashi OKADO et al.
; TITLE OF INVENTION: A GENE CODING FOR A PROTEIN REGULATING
; NUMBER OF SEQUENCES: 22
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Wenderoth, Lind & Ponack
; STREET: 805 Fifteenth Street, N.W., #700
; CITY: Washington
; STATE: D.C.

US-10-799-747-116 (1-20) x US-09-414-006-18 (1-1601)
Qy 4 HisSerValLeuSerPheLeuLeuTrpThrProTyrAlaLeu 17
Db 843 CATAGCATCCTTGATATTGGCTGGGTTCCATATGGAGTC 802

RESULT 17
US-09-414-006-19/c
; Sequence 19, Application US/09414006
; Patent No. 6348577
; GENERAL INFORMATION:
; APPLICANT: Takashi OKADO et al.
; TITLE OF INVENTION: A GENE CODING FOR A PROTEIN REGULATING
; NUMBER OF SEQUENCES: 22
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Wenderoth, Lind & Ponack
; STREET: 805 Fifteenth Street, N.W., #700
; CITY: Washington
; STATE: D.C.

US-10-799-747-116 (1-20) x US-09-414-006-18 (1-1601)
Qy 4 HisSerValLeuSerPheLeuLeuTrpThrProTyrAlaLeu 17
Db 843 CATAGCATCCTTGATATTGGCTGGGTTCCATATGGAGTC 802

RESULT 18
US-09-414-006-20/c
; Sequence 20, Application US/09414006
; Patent No. 6348577
; GENERAL INFORMATION:
; APPLICANT: Takashi OKADO et al.
; TITLE OF INVENTION: A GENE CODING FOR A PROTEIN REGULATING
; NUMBER OF SEQUENCES: 22
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Wenderoth, Lind & Ponack
; STREET: 805 Fifteenth Street, N.W., #700
; CITY: Washington
; STATE: D.C.

US-10-799-747-116 (1-20) x US-09-414-006-18 (1-1601)
Qy 4 HisSerValLeuSerPheLeuLeuTrpThrProTyrAlaLeu 17
Db 843 CATAGCATCCTTGATATTGGCTGGGTTCCATATGGAGTC 802


```

:
: COUNTRY: U.S.A.
: ZIP: 20005
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Diskette, 5.25 inch, 500 kb
: COMPUTER: IBM Compatible
: OPERATING SYSTEM: MS-DOS
: SOFTWARE: Wordperfect 5.1
:
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/09/447,223
: FILING DATE: 23-No. 6432664-1999
: CLASSIFICATION: <Unknown>
:
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: 08/243,403
: FILING DATE: May 16, 1994
: ATTORNEY/AGENT INFORMATION:
: NAME: Warren M. Cheek, Jr.
: REGISTRATION NUMBER: 33,367
: REFERENCE/DOCKET NUMBER: <Unknown>
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 202-371-8850
: TELEFAX: <Unknown>
: TELEX: <Unknown>
: INFORMATION FOR SEQ ID NO: 17:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 1601
: TYPE: nucleic acid
: STRANDEDNESS: double
: TOPOLOGY: linear
: MOLECULE TYPE: genomic DNA
: ANTI-SENSE: Yes
: SEQUENCE DESCRIPTION: SEQ ID NO: 17:
US-09-447-223-17

Alignment Scores:
Pred. No.: 68.2 Length: 1601
Score: 48.00 Matches: 7
Percent Similarity: 64.29% Conservative: 2
Best Local Similarity: 50.00% Mismatches: 5
Query Match: 48.48% Indels: 0
DB: 4 Gaps: 0

US-10-799-747-116 (1-20) X US-09-447-223-17 (1-1601)

Qy 4 HisSerValLeuSerPheLeuLeuTrpThrProTyrAlaLeu 17
Db 843 CATAGCATCCTGTGATATTTGGCCTGGTTCATATGGAGTC 802

```

Search completed: July 21, 2004, 02:54:35
Job time : 75 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright. (c) 1993 - 2004 Compugen Ltd.

QM protein - nucleic search, using frame_plus_p2n model

Run on: July 21, 2004, 02:14:01 ; Search time 355 Seconds
(without alignments)
274.852 Million cell updates/sec

Title: US-10-799-747-116

Perfect score: 99
Sequence: 1 MAHSVLSFLWTPYALKSX 20

Scoring table:
BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 3191023 seqs, 2439312756 residues

Total number of hits satisfying chosen parameters: 6382046

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:

-MODE=frame+ p2n.model -DEV=xlp
-Q=/cgn2_1/USPTO.spool_p/US10799747/runat 19072004 161420 21560/app query.fasta_1.199
-DB=Published Applications NA -QFMT=fastap -SUFFIX=rnpb -MINMATCH=0.1
-LOOPCL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=bloms62
-TRANS-human40 cdi -LIST=45 -LOCALIGN=200 -THR_SCORE=pct -THR_MAX=100
-THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0
-MAXLEN=2000000000 -USER=US10799747 @CGN 1 1 723 @runat 19072004 161420 21560
-NCPU=6 -ICPU=3 -NO MAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100
-LONGLOG -DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5
-FCAPOP=6 -FCAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Published Applications NA:

1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:
2: /cgn2_6/ptodata/2/pubpna/PCT_NEW PUB.seq:
3: /cgn2_6/ptodata/2/pubpna/US06_NEW PUB.seq:
4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq:
5: /cgn2_6/ptodata/2/pubpna/US07_NEW PUB.seq:
6: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq:
7: /cgn2_6/ptodata/2/pubpna/US08_NEW PUB.seq:
8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq:
9: /cgn2_6/ptodata/2/pubpna/US09A_PUBCOMB.seq:
10: /cgn2_6/ptodata/2/pubpna/US09B_PUBCOMB.seq:
11: /cgn2_6/ptodata/2/pubpna/US09C_PUBCOMB.seq:
12: /cgn2_6/ptodata/2/pubpna/US09_NEW PUB.seq:
13: /cgn2_6/ptodata/2/pubpna/US09_NEW PUB.seq:
14: /cgn2_6/ptodata/2/pubpna/US10A_PUBCOMB.seq:
15: /cgn2_6/ptodata/2/pubpna/US10B_PUBCOMB.seq:
16: /cgn2_6/ptodata/2/pubpna/US10C_PUBCOMB.seq:
17: /cgn2_6/ptodata/2/pubpna/US10_NEW PUB.seq:
18: /cgn2_6/ptodata/2/pubpna/US60_NEW PUB.seq:
19: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
------------	-------	-------------	--------	----	-------------

1	98	99.0	1434	15	US-10-195-730-12	Sequence 12, Appl
2	57	57.6	87731	13	US-10-087-192-1342	Sequence 1342, Ap
3	56	56.6	8417	16	US-10-074-024-635	Sequence 635, App
4	53	53.5	5760	16	US-10-074-024-426	Sequence 426, App
5	53	53.5	5763	16	US-10-074-024-425	Sequence 425, App
6	51	51.5	2500	15	US-10-364-360-20	Sequence 20, Appl
7	51	51.5	9641	15	US-10-364-360-22	Sequence 22, Appl
8	51	51.5	56577	13	US-10-087-192-1396	Sequence 1396, Ap
9	51	51.5	2940917	13	US-10-027-632-174763	Sequence 174763, Ap
10	51	51.5	2940917	16	US-10-027-632-174763	Sequence 174763, Ap
11	50	50.5	620	12	US-10-152-319A-215	Sequence 215, App
12	50	50.5	740	13	US-10-027-632-11891	Sequence 11891, A
13	50	50.5	740	16	US-10-027-632-11891	Sequence 11891, A
14	50	50.5	1206	13	US-10-027-632-250731	Sequence 250731, A
15	50	50.5	1206	13	US-10-027-632-250731	Sequence 250731, A
16	50	50.5	1206	16	US-10-027-632-250731	Sequence 250731, A
17	50	50.5	1206	16	US-10-027-632-250731	Sequence 250731, A
18	50	50.5	3034	17	US-10-437-963-7778	Sequence 7778, Ap
19	49	49.5	519	13	US-10-027-632-284792	Sequence 284792, Ap
20	49	49.5	519	16	US-10-027-632-284792	Sequence 284792, Ap
21	49	49.5	793	13	US-10-027-632-144194	Sequence 144194, A
22	49	49.5	793	13	US-10-027-632-144194	Sequence 144194, A
23	49	49.5	793	16	US-10-027-632-144194	Sequence 144194, A
24	49	49.5	793	16	US-10-027-632-144194	Sequence 144194, A
25	49	49.5	808	13	US-10-027-632-155999	Sequence 155999, A
26	49	49.5	808	16	US-10-027-632-155999	Sequence 155999, A
27	49	49.5	828	13	US-10-027-632-158555	Sequence 158555, A
28	49	49.5	828	13	US-10-027-632-158555	Sequence 158555, A
29	49	49.5	828	16	US-10-027-632-158555	Sequence 158555, A
30	49	49.5	828	16	US-10-027-632-158555	Sequence 158555, A
31	49	49.5	1092	17	US-10-437-963-61207	Sequence 61207, A
32	49	49.5	2393	13	US-10-027-632-102504	Sequence 102504, A
33	49	49.5	2393	16	US-10-027-632-102504	Sequence 102504, A
34	49	49.5	32768	9	US-09-070-927A-128	Sequence 128, App
35	48.5	49.0	72705	13	US-10-087-192-1966	Sequence 1966, Ap
36	48	48.5	290	14	US-10-040-739-637	Sequence 637, App
37	48	48.5	375	10	US-09-803-719-2326	Sequence 2326, Ap
38	48	48.5	495	12	US-09-969-034-1082	Sequence 1082, Ap
39	48	48.5	587	13	US-10-424-599-90540	Sequence 90540, A
40	48	48.5	612	14	US-10-002-344A-83	Sequence 83, Appl
41	48	48.5	802	13	US-10-424-599-111198	Sequence 111198, A
42	48	48.5	940	15	US-10-017-161-1223	Sequence 1223, Ap
43	48	48.5	940	16	US-10-292-798-1027	Sequence 1027, Ap
44	48	48.5	960	9	US-09-886-055-148	Sequence 148, App
45	48	48.5	960	10	US-09-804-291-148	Sequence 148, App

ALIGNMENTS

RESULT 1
US-10-195-730-12
; Sequence 12, Application US/10195730
; Publication No. US2003014492A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et. al
; TITLE OF INVENTION: 101 Human Secreted Proteins
; FILE REFERENCE: P2017P1
; CURRENT APPLICATION NUMBER: US/10/195,730
; CURRENT FILING DATE: 2002-07-16
; PRIOR APPLICATION NUMBER: US/09/281,976
; PRIOR FILING DATE: 1999-03-31
; PRIOR APPLICATION NUMBER: 60/060,837
; PRIOR FILING DATE: 1997-10-02
; PRIOR APPLICATION NUMBER: 60/060,862
; PRIOR FILING DATE: 1997-10-02
; NUMBER OF SEQ ID NOS: 390
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 12
; LENGTH: 1434
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-195-730-12

```
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc_feature
; LOCATION: (2140)
; OTHER INFORMATION: n equals a,t,g, or c
US-10-074-024-635
Alignment Scores:
Pred. No.: 75.7 Length: 8417
Score: 56.00 Matches: 9
Percent Similarity: 78.57% Conservatives: 2
Best Local Similarity: 64.29% Mismatches: 3
Query Match: 56.57% Indels: 0
DB: 16 Gaps: 0
US-10-799-747-116 (1-20) x US-10-074-024-635 (1-8417)
Qy 1 MetAlaAlaHisSerValLeuSerPheLeuLeuTrrProTyrAlaLeu 14
Db 934 ATGCCACACATTCGATTTTACGCTTTTATTTCTGGACCCCA 975
RESULT 4
US-10-074-024-426
; Sequence 426, Application US/10074024
; Publication No. US20030232975A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PC001C1
; CURRENT APPLICATION NUMBER: US/10/074,024
; CURRENT FILING DATE: 2002-02-14
; Prior Application removed - See file Wrapper or Palm
; NUMBER OF SEQ ID NOS: 879
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 426
; LENGTH: 5760
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-074-024-426
Alignment Scores:
Pred. No.: 159 Length: 5760
Score: 53.00 Matches: 10
Percent Similarity: 68.75% Conservatives: 1
Best Local Similarity: 62.50% Mismatches: 5
Query Match: 53.54% Indels: 0
DB: 16 Gaps: 0
US-10-799-747-116 (1-20) x US-10-074-024-426 (1-5760)
Qy 2 AlaAlaHisSerValLeuSerPheLeuLeuTrrProTyrAlaLeu 17
Db 567 GCCCGGCACTCAACCTTGAGCTTTGGCTTGGTTGATATTTCTT 614
RESULT 5
US-10-074-024-425
; Sequence 425, Application US/10074024
; Publication No. US20030232975A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PC001C1
; CURRENT APPLICATION NUMBER: US/10/074,024
; CURRENT FILING DATE: 2002-02-14
; Prior Application removed - See file Wrapper or Palm
; NUMBER OF SEQ ID NOS: 879
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 425
; LENGTH: 5763
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1871)
```

```
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc_feature
; LOCATION: (2140)
; OTHER INFORMATION: n equals a,t,g, or c
US-10-074-024-635
Alignment Scores:
Pred. No.: 1434 Length: 8417
Score: 56.00 Matches: 9
Percent Similarity: 78.57% Conservatives: 2
Best Local Similarity: 64.29% Mismatches: 3
Query Match: 56.57% Indels: 0
DB: 16 Gaps: 0
US-10-799-747-116 (1-20) x US-10-074-024-635 (1-8417)
Qy 1 MetAlaAlaHisSerValLeuSerPheLeuLeuTrrProTyrAlaLeu 14
Db 934 ATGCCACACATTCGATTTTACGCTTTTATTTCTGGACCCCA 975
RESULT 4
US-10-074-024-426
; Sequence 426, Application US/10074024
; Publication No. US20030232975A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PC001C1
; CURRENT APPLICATION NUMBER: US/10/074,024
; CURRENT FILING DATE: 2002-02-14
; Prior Application removed - See file Wrapper or Palm
; NUMBER OF SEQ ID NOS: 879
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 426
; LENGTH: 5760
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-074-024-426
Alignment Scores:
Pred. No.: 159 Length: 5760
Score: 53.00 Matches: 10
Percent Similarity: 68.75% Conservatives: 1
Best Local Similarity: 62.50% Mismatches: 5
Query Match: 53.54% Indels: 0
DB: 16 Gaps: 0
US-10-799-747-116 (1-20) x US-10-074-024-426 (1-5760)
Qy 2 AlaAlaHisSerValLeuSerPheLeuLeuTrrProTyrAlaLeu 17
Db 567 GCCCGGCACTCAACCTTGAGCTTTGGCTTGGTTGATATTTCTT 614
RESULT 5
US-10-074-024-425
; Sequence 425, Application US/10074024
; Publication No. US20030232975A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PC001C1
; CURRENT APPLICATION NUMBER: US/10/074,024
; CURRENT FILING DATE: 2002-02-14
; Prior Application removed - See file Wrapper or Palm
; NUMBER OF SEQ ID NOS: 879
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 425
; LENGTH: 5763
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1871)
```

; OTHER INFORMATION: n equals a,t,g, or c
US-10-074-024-425

Alignment Scores: Length: 5763
Pred. No.: 159
Score: 53.00
Percent Similarity: 68.75%
Best Local Similarity: 62.50%
Query Match: 53.54%
DB: 16
Gaps: 0

US-10-799-747-116 (1-20) x US-10-074-024-425 (1-5763)

Qy 2 AlaAlaHisSerValLeuSerPheLeuLeuTrrProTyAlaLeu 17
Db 567 GCCCGGCACTCAACCCCTTGAGCTTTTGGCTTTGGTTGTATATCTCTT 614

RESULT 6

US-10-364-360-20/c
; Sequence 20, Application US/10364360
; Publication No. US20030180324A1
; GENERAL INFORMATION:
; APPLICANT: GUERTLER, Lutz Gerhard
; APPLICANT: HAUSER, Hans Peter
; APPLICANT: DONGMO DELOKO, Yvette Beatrice
; APPLICANT: ZEKENG, Leopold
; APPLICANT: KAPTUE, Lazare
; TITLE OF INVENTION: LENTIVIRUS FROM THE GROUP OF IMMUNODEFICIENCY VIRUSES OF DRILL MC
; FILE REFERENCE: 067595/0106
; CURRENT APPLICATION NUMBER: US/10/364,360
; PRIOR FILING DATE: 2003-02-12
; PRIOR FILING DATE: 2000-07-29
; PRIOR FILING DATE: 1999-08-03
; NUMBER OF SEQ ID NOS: 57
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 20
; LENGTH: 2500
; TYPE: DNA
; ORGANISM: SIV - viral
US-10-364-360-20

Alignment Scores: Length: 2500
Pred. No.: 133
Score: 51.00
Percent Similarity: 100.00%
Best Local Similarity: 72.73%
Query Match: 51.52%
DB: 15
Gaps: 0

US-10-799-747-116 (1-20) x US-10-364-360-20 (1-2500)

Qy 7 LeuSerPheLeuLeuTrrProTyAlaLeu 17
Db 1453 TTGAGCTTTATACCTTTGGTCCCTACTCCTTG 1421

RESULT 7

US-10-364-360-22/c
; Sequence 22, Application US/10364360
; Publication No. US20030180324A1
; GENERAL INFORMATION:
; APPLICANT: GUERTLER, Lutz Gerhard
; APPLICANT: HAUSER, Hans Peter
; APPLICANT: DONGMO DELOKO, Yvette Beatrice
; APPLICANT: ZEKENG, Leopold
; APPLICANT: KAPTUE, Lazare
; TITLE OF INVENTION: LENTIVIRUS FROM THE GROUP OF IMMUNODEFICIENCY VIRUSES OF DRILL MC
; FILE REFERENCE: 067595/0106
; CURRENT APPLICATION NUMBER: US/10/364,360
; PRIOR FILING DATE: 2003-02-12

; PRIOR APPLICATION NUMBER: US/09/625,972
; PRIOR FILING DATE: 2000-07-29
; PRIOR APPLICATION NUMBER: DE 199 36 003.0
; PRIOR FILING DATE: 1999-08-03

; NUMBER OF SEQ ID NOS: 57
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 22
; LENGTH: 9641
; TYPE: DNA
; ORGANISM: SIV - viral
US-10-364-360-22

Alignment Scores: Length: 9641
Pred. No.: 639
Score: 51.00
Percent Similarity: 100.00%
Best Local Similarity: 72.73%
Query Match: 51.52%
DB: 15
Gaps: 0

US-10-799-747-116 (1-20) x US-10-364-360-22 (1-9641)

Qy 7 LeuSerPheLeuLeuTrrProTyAlaLeu 17
Db 1453 TTGAGCTTTATACCTTTGGTCCCTACTCCTTG 1421

RESULT 8

US-10-087-192-1396/c
; Sequence 1396, Application US/10087192
; Publication No. US20020182586A1
; GENERAL INFORMATION:
; APPLICANT: MORRIS, David W.
; APPLICANT: ENGELHARD, Eric K.
; TITLE OF INVENTION: NOVEL COMPOSITIONS AND METHODS FOR
; FILE REFERENCE: 529452000122
; CURRENT APPLICATION NUMBER: US/10/087,192
; PRIOR FILING DATE: 2002-03-01
; PRIOR APPLICATION NUMBER: US 09/747,377
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: US 09/798,586
; PRIOR FILING DATE: 2001-03-02
; NUMBER OF SEQ ID NOS: 2059
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1396
; LENGTH: 56577
; TYPE: DNA
; ORGANISM: Homo sapiens
; NAME/KEY: misc_feature
; LOCATION: (1)...(56577)
; OTHER INFORMATION: n = A,T,C or G
US-10-087-192-1396

Alignment Scores: Length: 56577
Pred. No.: 5,03e+03
Score: 51.00
Percent Similarity: 61.11%
Best Local Similarity: 55.56%
Query Match: 51.52%
DB: 13
Gaps: 0

US-10-799-747-116 (1-20) x US-10-087-192-1396 (1-56577)

Qy 1 MetAlaAlaHisSerValLeuSerPheLeuLeuTrrProTyAlaLeu 18
Db 29892 ATGATAGCTACTGCACTCTTGAACCTCTTGACTCCCAAGCTGCTGGAA 29839

RESULT 9

US-10-027-632-174763/c
; Sequence 174763, Application US/10027632
; Publication No. US20020198371A1
; GENERAL INFORMATION:

```
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 174763
; LENGTH: 2940917
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)..(2940917)
; OTHER INFORMATION: n = A,T,C or G
US-10-027-632-174763

Alignment Scores:
Pred. No.: 4.83e+05 Length: 2940917
Score: 51.00 Matches: 7
Percent Similarity: 84.62% Conservative: 4
Best Local Similarity: 53.85% Mismatches: 2
Query Match: 51.52% Indels: 0
DB: 13 Gaps: 0

US-10-799-747-116 (1-20) x US-10-027-632-174763 (1-2940917)
Qy 3 AlahisSerValLeuSerPheLeuLeuTrpThrProTyr 15
Db 189168 AGTCACGCCATTTGTTTTTTTGGTGGTCCATTT 189130

RESULT 10
US-10-632-174763/c
; Sequence 174763, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
US-10-027-632-174763

Alignment Scores:
Pred. No.: 4.83e+05 Length: 2940917
Score: 51.00 Matches: 7
Percent Similarity: 84.62% Conservative: 4
Best Local Similarity: 53.85% Mismatches: 2
Query Match: 51.52% Indels: 0
DB: 13 Gaps: 0

US-10-799-747-116 (1-20) x US-10-027-632-174763 (1-2940917)
Qy 3 AlahisSerValLeuSerPheLeuLeuTrpThrProTyr 15
Db 189168 AGTCACGCCATTTGTTTTTTTGGTGGTCCATTT 189130

RESULT 11
US-10-152-319A-215/c
; Sequence 215, Application US/10152319A
; Publication No. US20040072160A1
; GENERAL INFORMATION:
; APPLICANT: Mendrick, Donna
; APPLICANT: Porter, Mark
; APPLICANT: Johnson, Kory
; APPLICANT: Higgs, Brandon
; APPLICANT: Castle, Arthur
; APPLICANT: Elashoff, Michael
; TITLE OF INVENTION: Molecular Toxicology Modeling
; FILE REFERENCE: 44921-5089-US
; CURRENT APPLICATION NUMBER: US/10/152,319A
; CURRENT FILING DATE: 2002-05-22
; PRIOR APPLICATION NUMBER: US 60/292,335
; PRIOR FILING DATE: 2001-05-22
; PRIOR APPLICATION NUMBER: US 60/297,523
; PRIOR FILING DATE: 2001-06-13
; PRIOR APPLICATION NUMBER: US 60/298,925
; PRIOR FILING DATE: 2001-06-19
; PRIOR APPLICATION NUMBER: US 60/303,810
; PRIOR FILING DATE: 2001-07-10
; PRIOR APPLICATION NUMBER: US 60/303,807
; PRIOR FILING DATE: 2001-07-10
; PRIOR APPLICATION NUMBER: US 60/303,808
; PRIOR FILING DATE: 2001-07-10
; PRIOR APPLICATION NUMBER: US 60/315,047
; PRIOR FILING DATE: 2001-08-28
; PRIOR APPLICATION NUMBER: US 60/324,928
; PRIOR FILING DATE: 2001-09-27
; PRIOR APPLICATION NUMBER: US 60/330,867
; PRIOR FILING DATE: 2001-11-01
; PRIOR APPLICATION NUMBER: US 60/330,462
; PRIOR FILING DATE: 2001-10-22
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 2221
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 215
; LENGTH: 620
; TYPE: DNA
; ORGANISM: Rattus norvegicus
; FEATURE:
; OTHER INFORMATION: Genbank Accession No. AA891812
; NAME/KEY: misc feature
; LOCATION: (1)..(620)
; OTHER INFORMATION: n = a o r c o r g o r t
US-10-152-319A-215
```


Best Local Similarity: 57.14% Mismatches: 4
Query Match: 50.51% Indels: 0
DB: 13 Gaps: 0
US-10-799-747-116 (1-20) x US-10-027-632-250731 (1-1206)
Qy 6 ValLeuSerPheLeuThrProTyrAlaLeuLysSer 19
Db 704 ATTTGAGCTTCCTATTCTGGCTACCATACCCCATCTCCAGC 745

RESULT 15
US-10-027-632-250732
; Sequence 250732, Application US/10027632
; Publication No. US20020198371A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; TITLE OF INVENTION: Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 250732
; LENGTH: 1206
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-250732

Alignment Scores:
Pred. No.: 84.1 Length: 1206
Score: 50.00 Matches: 8
Percent Similarity: 71.43% Conservative: 2
Best Local Similarity: 57.14% Mismatches: 4
Query Match: 50.51% Indels: 0
DB: 13 Gaps: 0
US-10-799-747-116 (1-20) x US-10-027-632-250732 (1-1206)
Qy 6 ValLeuSerPheLeuThrProTyrAlaLeuLysSer 19
Db 704 ATTTGAGCTTCCTATTCTGGCTACCATACCCCATCTCCAGC 745

Search completed: July 21, 2004, 04:13:52
Job time : 1132 secs

GenCore version 5.1.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: July 21, 2004, 01:17:25 ; Search time 2469 Seconds
(without alignments)
241.897 Million cell updates/sec

Title: US-10-799-747-116
Perfect score: 99

Sequence: 1 MAHSVLSFLWTPYALKSX 20

Scoring table: BLOSUM62

Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 27513289 seqs, 14931090276 residues

Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DEV=xlp
-Q=/cgn2_1/USPTO_spool_p/US10799747/runat 19072004 161416 21448/app_query.fasta.1.199
-DB=EST -QFMT=fastcap -SUFFIX=rst -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=-1 -NATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-LOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000
-USER=US10799747@cgn_1_15180 @runat 19072004 161416 21448 -NCPU=6 -ICPU=3
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

EST:
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_est1:*
10: gb_est2:*
11: gb_htc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: em_gss_hum:*
18: em_gss_inv:*
19: em_gss_pln:*
20: em_gss_vrt:*
21: em_gss_fun:*
22: em_gss_mam:*
23: em_gss_mus:*
24: em_gss_pro:*
25: em_gss_rod:*
26: em_gss_phg:*
27: em_gss_vrl:*
28: gb_gss1:*

29: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	98	99.0	240	10	BF910533 CM4-UT004
2	98	99.0	575	12	BM722991 UI-E-R01
3	98	99.0	765	12	BI914473 603182264
4	98	99.0	785	12	BG484396 602505037
5	98	99.0	1201	13	EX441923 BX441923
6	56	56.6	352	9	AA654972 nt61e09.s
C 7	55	55.6	642	29	CE789116 tigr-gss-
8	55	55.6	723	28	AZ831423 2M0111H12
C 9	54	54.5	550	29	TA226G11P
10	54	54.5	724	28	AQ645123 RPCI93-EC
11	54	54.5	816	14	CA967853 CCLX04a13
C 12	53	53.5	394	9	AA766276 oa238c04.s
C 13	53	53.5	429	9	AA804444 nz38c04.s
C 14	53	53.5	600	12	BI140807 IPI 39.B0
15	53	53.5	603	12	BM571235 fw75h06.x
16	53	53.5	635	13	BU262936 603818292
17	53	53.5	641	13	BU350638 603526071
18	53	53.5	641	13	BU448459 603211744
19	53	53.5	643	13	BU353991 603529626
20	53	53.5	643	13	BU394939 603511454
21	53	53.5	649	13	BU337269 603513967
22	53	53.5	657	13	BU330405 603512306
23	53	53.5	784	29	CG759218 ZMMBB5031
24	53	53.5	804	28	AZ906788 RPCI-24-1
25	53	53.5	904	29	CG768437 TCB42.3 F
C 26	52.5	53.0	348	28	CC390782 PUENH29TD
27	52	52.5	254	9	AA324608 EST27458
28	52	52.5	259	28	AQ104627 HS_3047_B
C 29	52	52.5	262	9	AI164098 A054p15u
30	52	52.5	288	14	F07932 HSC2LH031 n
C 31	52	52.5	293	10	BB564042 BB564042
C 32	52	52.5	372	29	CC717225 OGCLD72TV
33	52	52.5	374	9	AA411283 zv26f01.x
C 34	52	52.5	390	10	AW338878 ha58a11.x
C 35	52	52.5	423	28	CC370277 PUHGP23TD
C 36	52	52.5	584	9	AI754441 cr25d04.x
C 37	52	52.5	607	9	AI672090 ty63d04.x
C 38	52	52.5	653	28	CC004087 PUDH064TD
C 39	52	52.5	669	28	AZ572182 302PVC03
C 40	52	52.5	691	13	BU871645 Q032H01 P
C 41	52	52.5	705	14	CD218841 pgrin.pk0
C 42	52	52.5	749	29	CG296747 OG3C117TV
C 43	52	52.5	759	28	BI837121 603090085
C 44	52	52.5	759	28	BZ985367 PUECT180TD
45	52	52.5	837	29	CC717215 OGCLD72TH

ALIGNMENTS

RESULT 1
BF910533/c
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS

BF910533
CM4-UT0042-011100-395-f08 UT0042 Homo sapiens cDNA, mRNA sequence.
BF910533
BF910533.1 GI:12301991
EST.
Homo sapiens (human)
Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,

240 bp
mRNA
linear
EST 18-JAN-2001

TITLE
Shotgun sequencing of the human transcriptome with ORF expressed
sequence tags

JOURNAL
Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)

MEDLINE
20202663

PUBMED
10737800

COMMENT
Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome
Project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?tl=CM4&t2=CM4-UT0042-
011100-395-f08&t3=2000-11-01&t4=1)
Seq primer: puc 18 forward
High quality sequence start: 81
High quality sequence stop: 240.

FEATURES
source
1..240
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/dev_stage="Adult"
/clone_lib="UT0042"
/note="Organ: uterus_tumor; Vector: puc18; Site 1: SmaI;
Site 2: SmaI; A mini-library was made by cloning products
derived from ORESTES PCR (U.S. Letters Patent application
No. 196,716 - Ludwig Institute for Cancer Research)
profiles into the pUC 18 vector. Reverse transcription of
tissue mRNA and cDNA amplification were performed under
low stringency conditions."

ORIGIN
Alignment Scores:
Pred. No.: 8 23e-06 Length: 240
Score: 98.00 Matches: 19
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 98.99% Indels: 0
DB: 10 Gaps: 0
US-10-799-747-116 (1-20) x BF910533 (1-240)
QY 1 MetaAlaHisSerValLeuSerPheLeuThrProTyAlaLeuLysSer 19
DB 191 ATGGCAGCCCATTCAGTCTTGAGTTTCTTCTCTGGACACCTTATGCTCTGAATCA 135

ORIGIN
Alignment Scores:
Pred. No.: 2 54e-05 Length: 575
Score: 98.00 Matches: 19
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 98.99% Indels: 0
DB: 12 Gaps: 0
US-10-799-747-116 (1-20) x BM722991 (1-575)
QY 1 MetaAlaHisSerValLeuSerPheLeuThrProTyAlaLeuLysSer 19
DB 454 ATGGCAGCCCATTCAGTCTTGAGTTTCTTCTCTGGACACCTTATGCTCTGAATCA 510

RESULT 2
BM722991
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
MEDLINE

RESULT 3
BM722991
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS

BM722991 575 bp mRNA linear EST 01-MAR-2002
UI-E-E01-aid-a-14-0-UI.r1 UI-E-E01 Homo sapiens cDNA clone
UI-E-E01-aid-a-14-0-UI 5', mRNA sequence.
BM722991
EST.
Homo sapiens (human)
Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 575)
Ronaldo, M.F., Lennon, G. and Soares, M.B.
Normalization and subtraction: two approaches to facilitate gene
discovery
Genome Res. 6 (9), 791-806 (1996)
97044477

BM722991 765 bp mRNA linear EST 16-OCT-2001
603182264F1 NIH_MGC_121 Homo sapiens cDNA clone IMAGE:5246463 5',
mRNA sequence.
BI914473
EST.
Homo sapiens (human)
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 765)
NIH-MGC http://mgi.nci.nih.gov/.

8889548
Contact: Soares, MB
Coordinated Laboratory for Computational Genomics
University of Iowa
375 Newton Road , 4156 MEBRF, Iowa City, IA 52242, USA
Tel: 319 335 8250
Fax: 319 335 9565
Email: bento-soares@uiowa.edu
Tissue Procurement: Dr. Gregg Hageman
cDNA Library preparation: Dr. M. Bento Soares, University of Iowa
cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
Clone Distribution: Researchers may obtain clones from Research
Genetics (www.resgen.com).
The following repetitive elements were found in this cDNA
sequence: 148-169, >AT-richLow_complexity (matched complement)
Seq primer: M13 Reverse.
Location/Qualifiers
1..575
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="UI-E-E01-aid-a-14-0-UI"
/tissue_type="fetal eye"
/dev_stage="fetal"
/lab_host="DH10B (Life Technologies) (T1 phage resistant)"
/clone_lib="UI-E-E01"
/note="Organ: eye; Vector: pT7T3-Pac (Pharmacia) with a
modified polylinker; Site 1: EcoR I; Site 2: Not I;
UI-E-E01 is a normalized cDNA library containing the
following tissue(s): fetal eye. The library was
constructed according to Bonaldo, Lennon and Soares,
Genome Research, 6:791-806, 1996. First strand cDNA
synthesis was primed with an oligo-dT primer containing a
Not I site. Double stranded cDNA was ligated to an EcoR I
adaptor, digested with Not I, and cloned directionally
into pT7T3-Pac vector. The oligonucleotide used to prime
the synthesis of first-strand cDNA contains a library tag
sequence that is located between the Not I site and the
(dT)18 tail. The sequence tag for this library is
CGCGTATACC. This library was created for the program, Gene
Discovery in the Visual System, supported by National Eye
Institute (NEI)."

8889548
Contact: Soares, MB
Coordinated Laboratory for Computational Genomics
University of Iowa
375 Newton Road , 4156 MEBRF, Iowa City, IA 52242, USA
Tel: 319 335 8250
Fax: 319 335 9565
Email: bento-soares@uiowa.edu
Tissue Procurement: Dr. Gregg Hageman
cDNA Library preparation: Dr. M. Bento Soares, University of Iowa
cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
Clone Distribution: Researchers may obtain clones from Research
Genetics (www.resgen.com).
The following repetitive elements were found in this cDNA
sequence: 148-169, >AT-richLow_complexity (matched complement)
Seq primer: M13 Reverse.
Location/Qualifiers
1..575
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="UI-E-E01-aid-a-14-0-UI"
/tissue_type="fetal eye"
/dev_stage="fetal"
/lab_host="DH10B (Life Technologies) (T1 phage resistant)"
/clone_lib="UI-E-E01"
/note="Organ: eye; Vector: pT7T3-Pac (Pharmacia) with a
modified polylinker; Site 1: EcoR I; Site 2: Not I;
UI-E-E01 is a normalized cDNA library containing the
following tissue(s): fetal eye. The library was
constructed according to Bonaldo, Lennon and Soares,
Genome Research, 6:791-806, 1996. First strand cDNA
synthesis was primed with an oligo-dT primer containing a
Not I site. Double stranded cDNA was ligated to an EcoR I
adaptor, digested with Not I, and cloned directionally
into pT7T3-Pac vector. The oligonucleotide used to prime
the synthesis of first-strand cDNA contains a library tag
sequence that is located between the Not I site and the
(dT)18 tail. The sequence tag for this library is
CGCGTATACC. This library was created for the program, Gene
Discovery in the Visual System, supported by National Eye
Institute (NEI)."

TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: Life Technologies, Inc.
cDNA Library Preparation: Life Technologies, Inc.
DNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLCM11621 row: p column: 16
High quality sequence stop: 762.

FEATURES

source
Location/Qualifiers
1..765
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:5246463"
/lab_host="DH10B"
/clone_lib="NIH_MGC_121"

/note="Organ: brain; Vector: pCMV-SPORT6; Site 1: NotI; Site 2: EcoRV (destroyed); RNA source anonymous pool of 3 fetal brains, female age 20 weeks, female age 24 weeks, and male age 26 weeks. Library is oligo-dT primed and directionally cloned (EcoRV site is destroyed upon cloning). Average insert size 1.7 kb, insert size range 0.7-3.5 kb. Library is normalized and enriched for full-length clones and was constructed by C. Gruber (Invitrogen). Research Genetics tracking code 017. Note: this is a NIH_MGC Library."

ORIGIN

Alignment Scores:
Pred. No.: 3.67e-05 Length: 765
Score: 98.00 Matches: 19
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 98.99% Indels: 0
DB: 12 Gaps: 0

US-10-799-747-116 (1-20) x BI914473 (1-765)

QY 1 MetAlaAlaHisSerValLeuSerPheLeuLeuThrProTyrAlaLeuLysSer 19
DB 556 ATGGACCCCATTCAGTCTTGAGTTTCTTCTGGACACCTTATGCTCTGAAATCA 612

RESULT 4
BG484396 785 bp mRNA linear EST 21-MAR-2001
LOCUS 602505037F1 NIH_MGC_77 Homo sapiens cDNA clone IMAGE:4618473 5',
DEFINITION mRNA sequence.

ACCESSION BG484396 GI:13416675
VERSION BG484396
KEYWORDS EST.
SOURCE Homo sapiens (human)

ORGANISM

Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 785)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)

COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: CLONTECH Laboratories, Inc.
cDNA Library Preparation: CLONTECH Laboratories, Inc.
DNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLCM1376 row: j column: 10

High quality sequence start: 18
High quality sequence stop: 751.
Location/Qualifiers

FEATURES

source
Location/Qualifiers
1..785
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:4618473"
/lab_host="DH10B (T1 phage-resistant)"
/clone_lib="NIH_MGC_77"

/note="Organ: lung; Vector: pDNR-LIB (Clontech); Site 1: SfiI (ggcgcttcggcc); Site 2: SfiI (ggccattatggcc); 5' and 3' adaptors were used in cloning as follows: 5' adaptor sequence: 5'-CACGGCCATTATGGCC-3' and 3' adaptor sequence: 5'-ATTCTAGAGCGGAGGGCGGCATG-dt(30)BN-3' (where B = A, C, or G and N = A, C, G, or T). Average insert size 1.9 kb (range 0.5-4.0 kb). 12/15 colonies contained inserts by PCR. This library was enriched for full-length clones and was constructed by Clontech Laboratories (Palo Alto, CA). Note: this is a NIH_MGC Library."

ORIGIN

Alignment Scores:
Pred. No.: 3.79e-05 Length: 785
Score: 98.00 Matches: 19
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 98.99% Indels: 0
DB: 12 Gaps: 0

US-10-799-747-116 (1-20) x BG484396 (1-785)

QY 1 MetAlaAlaHisSerValLeuSerPheLeuLeuThrProTyrAlaLeuLysSer 19
DB 322 ATGGACCCCATTCAGTCTTGAGTTTCTTCTGGACACCTTATGCTCTGAAATCA 378

RESULT 5

BX441923 1201 bp mRNA linear EST 15-MAY-2003
LOCUS CS0DF023Y002 5-PRIME, mRNA sequence.
DEFINITION BX441923 Homo sapiens FETAL BRAIN Homo sapiens cDNA clone

ACCESSION BX441923 GI:30771989
VERSION BX441923
KEYWORDS EST.
SOURCE Homo sapiens (human)

ORGANISM

Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 1201)
AUTHORS Li, W.B., Gruber, C., Jessee, J., and Polayes, D.
TITLE Full-length cDNA libraries and normalization
JOURNAL Unpublished (2001)
COMMENT Contact: Genoscope

Genoscope - Centre National de Sequencage
BP 191 91006 EVRY cedex - France

Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr

Library was constructed by Life Technologies, a division of
Invitrogen. This sequence belongs to sequence cluster 2532.r For
more information about this cluster, see
http://www.genoscope.cns.fr/

cgi-bin/cluster.cgi?seq=CS0DF023BH01Q1&cluster=2532.r. Contact :
Feng Liang Email : fliang@lifetech.com URL :
http://fulllength.invitrogen.com/ Invitrogen Corporation 1600
Faraday Avenue Genoscope sequence ID : CS0DF023BH01Q1.

FEATURES

source
Location/Qualifiers
1..1201
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CS0DF023Y002"
/tissue_type="FETAL BRAIN"
/dev_stage="fetal"
/clone_lib="Homo sapiens FETAL BRAIN"

/note="Organ: brain; Vector: pCMVSPORT 6; 1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-strand cDNA was digested with Not I and cloned into the Not I and EcoRV sites of the pCMVSPORT 6 vector. Library was not normalized."

ORIGIN

Alignment Scores:

Pred. No.:	6.56e-05	Length:	1201
Score:	98.00	Matches:	19
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	98.99%	Indels:	0
DB:	13	Gaps:	0

US-10-799-747-116 (1-20) x BX441923 (1-1201)

Qy 1 MetAlaLaHisSerValLeuSerPheLeuLeuThrProTyrAlaLeuLysSer 19

Db 146 ATGGCAGCCCAATCAGCTTGAGTTTCTTCTTCGACACCTTATGCTCTGAAATCA 202

RESULT 6

AA654972

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

CONTACT: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: W. Marston Linehan, M.D., Rodrigo Chuaqui, M.D., Michael Emmert-Buck, M.D., Ph.D.
cDNA Library Preparation: David B. Krizman, Ph.D.
cDNA Library Arrayed by: Genome Systems Inc., Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/ILNL at: www-bio.llnl.gov/bbrp/image/image.html
Seq primer: -40ml3 fwd. ET from Amersham.

FEATURES

source

1..352

/organism="Homo sapiens"

/mol_type="mRNA"

/db_xref="taxon:9606"

/clone="IMAGE:1203016"

/sex="Male"

/dev_stage="45 years old"

/lab_host="DH10B"

/clone_lib="NCI CGAP Pr3"

/note="Vector: pAMP10; Site 1: NotI; Site 2: EcoRI; 1st strand cDNA was primed with oligo(dT)17 on 50 ng of DNase-treated, total cellular RNA obtained from 5,000-10,000 microdissected cells histologically determined to be fully malignant prostate cancer cells. Double-stranded cDNA was ligated to EcoRI adaptors, 5 cycles of PCR applied to the cDNA with an adaptor-specific primer, and the resulting PCR product subcloned into pAMP10 by the UDG-cloning method (Life Technologies). Average insert size is 600 bp. NOTE: Not directionally cloned. This library was constructed by David Krizman."

ORIGIN

Alignment Scores:

Pred. No.:	57.7	Length:	352
Score:	56.00	Matches:	9
Percent Similarity:	80.00%	Conservative:	3
Best Local Similarity:	60.00%	Mismatches:	3
Query Match:	56.57%	Indels:	0
DB:	9	Gaps:	0

US-10-799-747-116 (1-20) x AA654972 (1-352)

Qy 4 HisSerValLeuSerPheLeuLeuThrProTyrAlaLeuLys 18

Db 11 CACAGCAAAATATCATTCATTATTGGTGGCCACAGCCCAATTAAG 55

RESULT 7

CE789116/c

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

COMMENT

CONTACT: Kirkness EP
The Institute for Genomic Research
Department of Eukaryotic Genomics, TIGR, 9712 Medical Center Drive, Rockville, MD 20850, USA
Tel: 301-838-0200
Fax: 301-838-0208
Email: ekirkness@tigr.org
Class: shotgun.

FEATURES

source

1..642

/organism="Canis familiaris"

/mol_type="genomic DNA"

/strain="Standard Poodle"

/db_xref="taxon:9615"

/clone_lib="Dog Library"

/note="Site 1: BatXi; Libraries were prepared from peripheral blood"

ORIGIN

Alignment Scores:

Pred. No.:	180	Length:	642
Score:	55.00	Matches:	10
Percent Similarity:	84.62%	Conservative:	1
Best Local Similarity:	76.92%	Mismatches:	2
Query Match:	55.56%	Indels:	0
DB:	29	Gaps:	0

US-10-799-747-116 (1-20) x CE789116 (1-642)

Qy 7 LeuSerPheLeuLeuThrProTyrAlaLeuLysSer 19

Db 208 TTAATCATTCCTTCCTTGGTTCCTATTTATTAAGC 170

RESULT 8

AZ831423

LOCUS

DEFINITION

ACCESSION

2M0111H12F Mouse 10kb plasmid UUGC1M library Mus musculus genomic clone UUGC2M0111H12 F, genomic survey sequence.

```

VERSION      AZB31423.1  GI:13001331
KEYWORDS     Mus musculus (house mouse)
SOURCE       Mus musculus
ORGANISM     Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE    1 (bases 1 to 723)
AUTHORS      Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
            Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T.,
            Reilly,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von
            Niederhausern,A. and Wright,D., Weiss,R.
TITLE        Mouse whole genome scaffolding with paired end reads from 10kb
            plasmid inserts
JOURNAL      Unpublished (2000)
COMMENT      Contact: Robert B. Weiss
            University of Utah Genome Center
            Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
            84112, USA
            Tel: 801 585 5606
            Fax: 801 585 7177
            Email: ddunn@genetics.utah.edu
            Insert Length: 10000 Std Error: 0.00
            Plate: 0111 row: H column: 12
            Seq primer: CGTTGTAACAGCGGCAGT
            Class: plasmid ends
            High quality sequence stop: 723.

FEATURES     source
             1..723
                /organism="Mus musculus"
                /mol_type="genomic DNA"
                /strain="C57BL/6J"
                /db_xref="taxon:10090"
                /clone="UUGC2M0111H12"
                /sex="Male"
                /lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
                /clone_lib="Mouse 10kb plasmid UUGC1M library"
                /note="Vector: pWD42nv; Purified genomic DNA from M.
            Laboratory Mouse DNA Resource
            (http://www.jax.org/resources/documents/dnares/). The DNA
            was hydrodynamically sheared by repeated passage through a
            0.005 inch orifice at constant velocity. The sheared DNA
            was blunt end-repaired with T4 DNA polymerase and T4
            polynucleotide kinase. Adaptor oligonucleotides were
            ligated to the blunt ends in high molar excess. The
            adaptor DNA was purified and size-selected for a 9.5 to
            10.5 kb range using preparative agarose gel
            electrophoresis. Vector DNA was prepared from a derivative
            of pWD42 (gil4732114|gb|AF129072.1), a copy-number
            inducible derivative of plasmid R1. The vector was ligated
            with adaptors complementary to the insert adaptors and
            purified. The sheared, adaptor mouse DNA was annealed to
            adaptor vector DNA, and transformed into
            chemically-competent E. coli XL10-Gold (Stratagene) cells
            and selected for ampicillin resistance."

ORIGIN
Alignment Scores:      210      Length:      723
Pred. No.:            55.00      Matches:      9
Score:                66.67%      Conservative:  3
Percent Similarity:    50.00%      Mismatches:   6
Best Local Similarity: 55.56%      Indels:       0
Query Match:          28          Gaps:         0
DB:

US-10-799-747-116 (1-20) x AZB31423 (1-723)

QY      1 MetAlaHisSerValLeuSerPheLeuLeuTrrProTyrAlaLeuTys 18
      |||:::|||||
Db      121 ATGCTCTCTCACAGCCTTGCAATTACTTGTGTGTTACCTTATCCCTTAG 174
      |||:::|||||

RESULT 9

```

```

TA226G11P/c
LOCUS       TA226G11P      550 bp      DNA      linear      GSS 13-DEC-2000
DEFINITION  T. brucei sheared genomic DNA clone 226g11, forward sequence,
            genomic survey sequence.
ACCESSION   AL480086
VERSION     AL480086.1  GI:11845965
KEYWORDS    GSS.
SOURCE      Trypanosoma brucei
ORGANISM    Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae;
            Trypanosoma.
REFERENCE   1 (bases 1 to 550)
AUTHORS     Hall,N., Bowman,S., Lennard,N.J., Doggett,J., Atkin,R.,
            Chillingworth,C., Ormond,D., Harris,B., El-Sayed,N., Hou,L.,
            Melville,S.E., Rajandream,M.A. and Barrell,B.G.
TITLE       Direct Submission
JOURNAL     Submitted (10-DEC-2000) Trypanosoma brucei genome sequencing
            project, Sanger Centre, The Wellcome Trust Genome Campus, Hinxton,
            Cambridge CB10 1SA, E-mail: barrell@sanger.ac.uk and
            nh@sanger.ac.uk
COMMENT     Constructed at the Institute for Genomic Research (TIGR),
            Rockville, MD. Genomic DNA isolated from a cloned population of
            Trypanosoma brucei (TREU927/4 Gurat 10.1) was mechanically sheared
            to give a tight size distribution (
            4 kb). The v + i method used for the library construction is
            described in detail in Smith, H. and Venter, J.C. (Making small
            insert libraries for whole genome shotgun sequencing projects. In
            Genome Sequencing: A Practical Approach, eds. M. Vaudin and B.
            Barrell, Oxford University Press, 1999).
            Email: nh@sanger.ac.uk
            Details of T. brucei sequencing at the Sanger Centre are available
            at http://www.sanger.ac.uk/Projects/T_brucei/.

FEATURES     source
             1..550
                /organism="Trypanosoma brucei"
                /mol_type="genomic DNA"
                /strain="TREU927"
                /db_xref="taxon:5691"
                /clone="226g11"

ORIGIN
Alignment Scores:      212      Length:      550
Pred. No.:            54.00      Matches:      11
Score:                73.33%      Conservative:  0
Percent Similarity:    73.33%      Mismatches:   4
Best Local Similarity: 54.55%      Indels:       0
Query Match:          29          Gaps:         0
DB:

US-10-799-747-116 (1-20) x TA226G11P (1-550)

QY      3 AlaHisSerValLeuSerPheLeuLeuTrrProTyrAlaLeu 17
      |||:::|||||
Db      356 GCGGGTGGTGTGCTTGTAGTTTTTTTGTGACCTTTAGCGCTG 312
      |||:::|||||

RESULT 10
AQ645123
LOCUS       AQ645123      724 bp      DNA      linear      GSS 08-JUL-1999
DEFINITION  RPC193-EcoRI-2M16-TV RPC193-EcoRI Trypanosoma brucei genomic clone
            genomic survey sequence.
ACCESSION   AQ645123
VERSION     AQ645123.1  GI:5121833
KEYWORDS    GSS.
SOURCE      Trypanosoma brucei
ORGANISM    Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae;
            Trypanosoma.
REFERENCE   1 (bases 1 to 724)
AUTHORS     El-Sayed,N., Zhao,S., Zhao,H., Gill,S., Suh,E., Malek,J., Fujii,C.,
            Gerrard,C., Leech,V., de Jong,P., Ullu,E., Melville,S.,
            Donelson,J., Fraser,C. and Adams,M.
            Use of BAC end sequences from Trypanosoma brucei GUTat 10.1 RPCI-93
            Library for gene discovery and sequence-ready map construction

ORIGIN

```

JOURNAL
COMMENT

Unpublished (1999)
Other GSs: RPC193-EcoRI-2M16.TP
Contact: Najib M. El-Sayed
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: nelsayed@tigr.org
Clones and high density filters may be purchased from BACPAC
Resources (http://bacpac.med.buffalo.edu). BAC end sequences search
page: http://www.tigr.org/tdb/mdb/tbdb/.

Seq primer: T7

Class: BAC ends.

FEATURES
source

Location/Qualifiers
1. .724
/organism="Trypanosoma brucei"
/mol_type="genomic DNA"
/strain="TREU927/4 GUTat 10.1"
/db_xref="taxon:5691"
/clone="RPC193-EcoRI-2M16"
/note="Vector: pBACe3.6; Site 1: Eco RI; Site 2: Eco RI;
Constructed for The Institute for Genomic Research by
Bohui Zhao in Pieter de Jong's laboratory (Roswell Park
Cancer Institute, Buffalo, NY). Briefly, Trypanosoma
brucei TREU927/4 GUTat 10.1 agarose embedded DNA was
partially digested with a combination of Eco RI and Eco RI
methylase (RPC193-EcoRI segment) or Dpn II (RPC193-DpnII
segment). High molecular weight fragments were ligated in
pBACe3.6 vector digested with Eco RI or Bam HI,
respectively. The average insert size is 141 Kb. Total
coverage (both segments): > 90 X the haploid
non-minichromosomal genome."

ORIGIN

Alignment Scores:
Pred. No.: 302 Length: 724
Score: 54.00 Matches: 11
Percent Similarity: 73.33% Conservative: 0
Best Local Similarity: 73.33% Mismatches: 4
Query Match: 54.55% Indels: 0
DB: 28 Gaps: 0

US-10-799-747-116 (1-20) x AQ645123 (1-724)

Qy 3 AlahisSerValLeuSerPheLeuTrpThrProTyrAlaLeu 17

Db 329 GCGGTTGGTGTCTAGTTTGTGGTCTCTTTAGCGCTG 373

RESULT 11

CA967853 816 bp mRNA linear EST 03-JAN-2003
LOCUS CcLX04a13k03f1 Carp mixed tissue library 1 Cyprinus carpio cDNA
DEFINITION clone 13k03 5', mRNA sequence.

ACCESSION CA967853

VERSION CA967853.1

KEYWORDS EST.

SOURCE Cyprinus carpio (common carp)

ORGANISM

Cyprinus carpio
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
Cypriniformes; Cyprinidae; Cyprinus.

1 (bases 1 to 816)

Gracey, A.Y., Fraser, E., Li, W., and Cossins, A.R.

Microarray and EST analysis of the carp (Cyprinus carpio)

transcriptome during environmental stress

Unpublished (2003)

JOURNAL

COMMENT Contact: Andrew R. Cossins

Laboratory for Environmental Gene Regulation

University of Liverpool

School of Biological Sciences, The Biosciences Building, Crown

Street, Liverpool, United Kingdom, L69 7ZB

Tel: +44 (0)151-795-4510
Fax: +44 (0)151-795-4431

Email: cossins@liverpool.ac.uk

Vector has been trimmed from this EST.

Plate: 13 row: k column: 03

Seq primer: Triplex 5' LD (5'-CTCGGAGCGCGCCATTGTGTGT-3')

High quality sequence start: 33

High quality sequence stop: 494.

FEATURES
source

Location/Qualifiers

1. .816

/organism="Cyprinus carpio"

/mol_type="mRNA"

/db_xref="taxon:7962"

/clone="13k03"

/sex="Male & female"

/tissue_type="Skeletal white muscle, cardiac muscle,

kidney, brain, gill, intestinal mucosa"

/dev_stage="Adult"

/lab_host="E.coli Electromax DH10B"

/clone_lib="Carp mixed tissue library 1"

/note="Vector: pTriplex2; Site 1: SfiI GGCCATTACGGCC;

Site 2: SfiI GGCGCTCGGCC; Normalized cDNA library

prepared from mixed tissues of warm, cold and hypoxia

challenged animals"

ORIGIN

Alignment Scores:
Pred. No.: 353 Length: 816
Score: 54.00 Matches: 10
Percent Similarity: 80.00% Conservative: 2
Best Local Similarity: 66.67% Mismatches: 3
Query Match: 54.55% Indels: 0
DB: 14 Gaps: 0

US-10-799-747-116 (1-20) x CA967853 (1-816)

Qy 5 SerValLeuSerPheLeuTrpThrProTyrAlaLeuLysSer 19

Db 679 TCTGCTCTCTTTCTTACTTTGGTTCCTTCCTCTTCTTC 723

RESULT 12

AA766276/c

LOCUS

DEFINITION

AA766276

oa29c04.s1 NCI CGAP GCB1 Homo sapiens cDNA

similar to contains element OFR repetitive element ;

sequence.

ACCESSION AA766276

VERSION AA766276.1

KEYWORDS EST.

SOURCE Homo sapiens (human)

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 394)

NCI-CCAP http://www.ncbi.nlm.nih.gov/ncicgap.

National Cancer Institute, Cancer Genome Anatomy Project (CGAP),

Tumor Gene Index

Unpublished (1997)

CONTACT: Robert Strausberg, Ph.D.

Email: ccapbs-remail.nih.gov

Tissue Procurement: Louis M. Staudt, M.D., Ph.D., David Allman,

Ph.D., Gerald Marti, M.D.

cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima

Bonaldo, Ph.D.

cDNA Library Arrayed by: Greg Lennon, Ph.D.

DNA Sequencing by: Washington University Genome Sequencing Center

Clone distribution: NCI-CCAP clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

www-bio.llnl.gov/bbrp/image/image.html

Insert Length: 666 Std Error: 0.00

Seq primer: -40ml3 fwd. ET from Amersham

High quality sequence stop: 357.

Location/Qualifiers

FEATURES

```

source
1. .394
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:1306374"
/tissue_type="germinal center B cell"
/lab_host="DH10B"
/clone_lib="NCI_CGAP_GCB1"
/notes="Vector: p7T3D-Pac (Pharmacia) with a modified
polylinker; Site 1: Not I; Site 2: Eco RI; 1st strand cDNA
was prepared from human tonsillar cells enriched for
germinal center B cells by flow sorting (CD20+, Igu-),
provided by Dr. Louis M. Staudt (NCI), Dr. David Allman
(NCI) and Dr. Gerald Marti (CBER). cDNA synthesis was
primed with a Not I - oligo(dT) primer
[5'-TGTTACCAATCTCAAGTGGGAGCGGCGCTCATTTTTTTTTTTT-3'
]. Double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not I
and Eco RI sites of the modified p7T3 vector. Library
went through one round of normalization, and was
constructed by Bento Soares and M. Fatima Bonaldo."

ORIGIN
Alignment Scores:
Pred. No.: 199 Length: 394
Score: 53.00 Matches: 9
Percent Similarity: 80.00% Conservative: 3
Best Local Similarity: 60.00% Mismatches: 3
Query Match: 53.54% Indels: 0
DB: 9 Gaps: 0

US-10-799-747-116 (1-20) x AA766276 (1-394)

Qy 2 AlaAlaHisSerValLeuSerPheLeuLeuTrpThrProTyrAla 16
|||||
Db 166 GCTGCACACTCTCTCTCTTTTATTCTCTGGCCCAATTTC 122

RESULT 13
AA804444/c
LOCUS
DEFINITION n238c04.sl NCI CGAP GCB1 Homo sapiens cDNA clone IMAGE:1290054 3',
similar to contains element QFR repetitive element 1, mRNA
sequence.
AA804444 429 bp mRNA linear EST 18-FEB-1998
VERSION AA804444.1 GI:2875957
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 429)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgaps-remail.nih.gov
Tissue Procurement: Louis M. Staudt, M.D., Ph.D., David Allman,
Ph.D., Gerald Marti, M.D.
CDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
Bonaldo, Ph.D.
CDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html
Insert Length: 649 Std Error: 0.00
Seq primer: -40ml3 fwd. ET from Amersham
High quality sequence stop: 249.
Location/Qualifiers
1. .429
/organism="Homo sapiens"
/mol_type="mRNA"

/db_xref="taxon:9606"
/clone="IMAGE:1290054"
/tissue_type="germinal center B cell"
/lab_host="DH10B"
/clone_lib="NCI_CGAP_GCB1"
/notes="Vector: p7T3D-Pac (Pharmacia) with a modified
polylinker; Site 1: Not I; Site 2: Eco RI; 1st strand cDNA
was prepared from human tonsillar cells enriched for
germinal center B cells by flow sorting (CD20+, Igu-),
provided by Dr. Louis M. Staudt (NCI), Dr. David Allman
(NCI) and Dr. Gerald Marti (CBER). cDNA synthesis was
primed with a Not I - oligo(dT) primer
[5'-TGTTACCAATCTCAAGTGGGAGCGGCGCTCATTTTTTTTTTTT-3'
]. Double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not I
and Eco RI sites of the modified p7T3 vector. Library
went through one round of normalization, and was
constructed by Bento Soares and M. Fatima Bonaldo."

ORIGIN
Alignment Scores:
Pred. No.: 222 Length: 429
Score: 53.00 Matches: 9
Percent Similarity: 80.00% Conservative: 3
Best Local Similarity: 60.00% Mismatches: 3
Query Match: 53.54% Indels: 0
DB: 9 Gaps: 0

US-10-799-747-116 (1-20) x AA804444 (1-429)

Qy 2 AlaAlaHisSerValLeuSerPheLeuLeuTrpThrProTyrAla 16
|||||
Db 166 GCTGCACACTCTCTCTTTTATTCTCTGGCCCAATTTC 122

RESULT 14
BI140807/c
LOCUS
DEFINITION IP1.39 B03.b1 A002 Immature pannicle 1 (IP1) Sorghum bicolor cDNA,
mRNA sequence.
BI140807 600 bp mRNA linear EST 03-JUL-2001
VERSION BI140807.1 GI:14593250
KEYWORDS EST.
SOURCE Sorghum bicolor (sorghum)
ORGANISM Sorghum bicolor
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
clade; Panicoideae; Andropogoneae; Sorghum.
1 (bases 1 to 600)
Klein,R.R., Cordonnier-Pratt,M.-M., Gingle,A., Sudman,M. and
Pratt,L.H.
An EST database from Sorghum: developing preanthesis pannicles
Unpublished (2001)
Contact: Cordonnier-Pratt MM
Laboratory for Genomics and Bioinformatics
The University of Georgia, Department of Plant Biology
Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA
Tel: 706 542 1860
Fax: 706 583 0210
Email: mmpratt@uga.edu
Sequences have been trimmed to exclude PolyA, vector and regions
below Phred quality 16. The threshold for high quality sequence is
20. Three-prime sequences, which are obtained with PolyTMix or T7
sequencing primer, are presented as the reverse complement.
High quality sequence stop: 538
POLYA-No. Location/Qualifiers
1. .600
/organism="Sorghum bicolor"
/mol_type="mRNA"
/cultivar="BTx623"
/db_xref="taxon:4558"
/clone_lib="Immature pannicle 1 (IP1)"

```


GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: July 21, 2004, 02:54:41 ; Search time 1704 Seconds

(without alignments)

508.721 Million cell updates/sec

Title: US-10-799-747-116

Perfect score: 20

Sequence: 1 MAHSVLSFLLWTPYALKSX 20

Scoring table: OLIGO
Xgapop 60.0 , Xgapext 60.0
Ygapop 60.0 , Ygapext 60.0
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 3470272 seqs, 21671516995 residues

Word size: 1

Total number of hits satisfying chosen parameters: 6933152

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Command line parameters: -DEV=xlp
-MODEL=frame+p2n.model
-Q=/cn2_1/USPTO.spool_p/US10799747/runat_19072004_161442_21981/app_query.fasta_1.199
-DB=GenEmbl -QFWT=fastap -SURFIX=olig.rge -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=oligo -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR SCORE=quality -THR MIN=1 -ALIGN=15 -MODE=LOCAL -OUTFMF=ptc
-NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10799747@CGN 1 5600 @runat_19072004_161442_21981 -NCFU=6 -ICPU=3
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=60 -XGAPEXT=60 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=60 -YGAPEXT=60 -DELOP=6 -DELEXT=7

Database :

GenEmbl.*

- 1: gb.ba.*
- 2: gb.htg.*
- 3: gb.in.*
- 4: gb.ov.*
- 5: gb.pat.*
- 6: gb.ph.*
- 7: gb.pl.*
- 8: gb.pr.*
- 9: gb.ro.*
- 10: gb.sts.*
- 11: gb.sy.*
- 12: gb.un.*
- 13: gb.vi.*
- 14: gb.ba.*
- 15: em.ba.*
- 16: em.fun.*
- 17: em.hum.*
- 18: em.in.*
- 19: em.mu.*
- 20: em.om.*
- 21: em.or.*
- 22: em.ov.*
- 23: em.pat.*
- 24: em.ph.*
- 25: em.pl.*
- 26: em.ro.*
- 27: em.sts.*
- 28: em.un.*

- 29: em.vi.*
- 30: em.htg.hum.*
- 31: em.htg.inv.*
- 32: em.htg.other.*
- 33: em.htg.mus.*
- 34: em.htg.pln.*
- 35: em.htg.fod.*
- 36: em.htg.mam.*
- 37: em.htg.vrt.*
- 38: em.sy.*
- 39: em.bto.hum.*
- 40: em.bto.mus.*
- 41: em.htgo.other.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	19	95.0	1434	6	BD078422	BD078422 101 human
2	19	95.0	3452	9	BC033650	BC033650 Homo sapi
3	19	95.0	3522	9	HSM804706	AL833393 Homo sapi
4	19	95.0	175081	9	AC021151	AC021151 Homo sapi
5	19	95.0	281662	2	AC117374	AC117374 Homo sapi
6	13	65.0	2020	9	HSM800227	AL049442 Homo sapi
7	9	45.0	153464	2	AC123991	AC123991 Mus muscu
8	9	45.0	167013	9	HSDJ828H9	AL121757 Human DNA
9	9	45.0	205725	10	AC122809	AC122809 Mus muscu
10	9	45.0	231768	2	AC110315	AC110315 Rattus no
11	9	45.0	254839	2	AC103176	AC103176 Rattus no
12	9	45.0	267352	2	AC098074	AC098074 Rattus no
13	9	45.0	306660	2	AC098226	AC098226 Rattus no
14	8	40.0	339	11	G25673	G25673 human STS E
15	8	40.0	356	11	G27758	G27758 human STS S
16	8	40.0	691	3	AK112944	AK112944 Clona int
17	8	40.0	1503	6	AX654294	AX654294 Sequence
18	8	40.0	1524	9	AB070060	AB070060 Macaca fa
19	8	40.0	1751	9	AK130941	AK130941 Homo sapi
20	8	40.0	1791	8	AK106862	AK106862 Oryza sat
21	8	40.0	1806	10	MMY15799	Y15799 Mus musculu
22	8	40.0	1858	10	MMY15797	Y15797 Mus musculu
23	24	8	1916	10	MMY15800	Y15800 Mus musculu
24	8	40.0	1952	10	AF040749	AF040749 Mus muscu
25	8	40.0	1975	8	AY090039	AY090039 Nicotiana
26	8	40.0	1997	10	AF040747	AF040747 Mus muscu
27	8	40.0	1999	10	AF040748	AF040748 Mus muscu
28	8	40.0	2841	10	MMY15798	Y15798 Mus musculu
29	8	40.0	3200	8	AK119280	AK119280 Oryza sat
30	8	40.0	3628	8	AF238476	AF238476 Oryza sat
31	8	40.0	14782	9	HSFLADPB	X02228 Human HLA-D
32	8	40.0	37201	9	AP001215	AP001215 Homo sapi
33	8	40.0	52058	2	AC100335	AC100335 Mus muscu
34	8	40.0	56083	2	AC141162	AC141162 Rattus no
35	35	8	60504	2	AC090716	AC090716 Homo sapi
36	8	40.0	57630	2	AC100885	AC100885 Mus muscu
37	8	40.0	62263	2	AC100239	AC100239 Mus muscu
38	8	40.0	63139	10	AL833790	AL833790 Mouse DNA
39	8	40.0	64380	9	BX120009	BX120009 Human DNA
40	8	40.0	65209	2	AC101355	AC101355 Mus muscu
41	8	40.0	67358	2	AC108426	AC108426 Mus muscu
42	8	40.0	68790	9	AF017732	AF017732 Homo sapi
43	8	40.0	74658	8	OS243961	AJ243961 Oryza sat
44	8	40.0	91585	10	AL606661	AL606661 Mouse DNA
45	8	40.0	93842	9	AL845446	AL845446 Human DNA

ALIGNMENTS

```

BD078422      1434 bp      DNA      linear      PAT 27-AUG-2002
LOCUS         101 human secretory proteins.
DEFINITION   BD078422
ACCESSION    BD078422
VERSION      BD078422.1 GI:22624025
KEYWORDS     JP 2001519156-A/11.
SOURCE       Homo sapiens (human)
ORGANISM     Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 1434)
Strausberg,R.L., Feingold,E.A., Grouse,L.H., Stern,M., Schuler,G.D.,
Klausner,R.D., Collins,F.S., Wagner,L., Shenmen,C.M., Schuler,G.D.,
Altschul,S.F., Zeeberg,B., Buetow,K.H., Schaefer,C.F., Bhat,N.K.,
Hopkins,R.F., Jordan,H., Moore,T., Max,S.I., Wang,J., Hsieh,F.,
Datchenko,L., Marusina,K., Farmer,A.A., Rubin,G.M., Hong,L.,
Shapleton,M., Soares,M.B., Bonaldo,M.F., Casavant,T.L.,
Schetz,T.E., Brownstein,M.J., Utsid,T.B., Toshiyuki,S.,
Carinci,P., Prange,C., Raha,S.S., Loquellano,N.A., Peters,G.J.,
Abramson,R.D., Mullaly,S.J., Bosak,S.A., McWhan,P.J.,
McKernan,K.J., Malek,J.A., Gunaratne,P.H., Richards,S.,
Worley,K.C., Hale,S., Garcia,A.M., Gay,L.J., Hulyk,S.W.,
Villalon,D.K., Muzny,D.M., Sodergren,E.J., Lu,X., Gibbs,R.A.,
Fahy,J., Helton,E., Kettman,M., Madan,A., Young,A.C., Shevchenko,Y.,
Sanchez,A., Whiting,M., Madan,A., Young,A.C., Shevchenko,Y.,
Bouffard,G.G., Blakesley,R.W., Touchman,J.W., Green,E.D.,
Dickson,M.C., Rodriguez,A.C., Grimwood,J., Schmutz,J., Myers,R.M.,
Butterfield,Y.S., Krzywinski,M.I., Skalska,U., Smalius,D.E.,
Schnerch,A., Schein,J.E., Jones,S.J. and Marra,M.A.
Generation and initial analysis of more than 15,000 full-length
human and mouse cDNA sequences
Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
2238257
PUBMED
12477932
REFERENCE
2 (bases 1 to 3452)
Strausberg,R.
Direct Submission
Submitted (02-JUL-2002) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA
NIH-MGC Project URL: http://mgc.nci.nih.gov
Contact: MGC help desk
Email: cgapbs@mail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: National Institutes of Health Intramural
Sequencing Center (NISC),
Gaithersburg, Maryland;
Web site: http://www.nisc.nih.gov/
Contact: nisc.mgc@nih.gov
Akhter,N., Ayele,K., Beckstrom-Sternberg,S.M., Benjamin,B.,
Blakesley,R.W., Bouffard,G.G., Breen,K., Brinkley,C., Brooks,S.,
Dietrich,N.L., Granite,S., Guan,X., Gupta,J., Haghighi,P.,
Hansen,N., Ho,S.-L., Karlins,E., Kwong,P., Laric,P., Legaspi,R.,
Maduro,Q.L., Masiello,C., Maskeri,B., Mastrian,S.D., McCloskey,J.C.,
McDowell,J., Pearson,R., Stantripop,S., Thomas,P.J., Touchman,J.W.,
Turgeon,C., Vogt,J.D., Walker,M.A., Wetherby,K.D., Wiggins,L.,
Young,A., Zhang,L.-H. and Green,E.D.
Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
Series: IRAK Plate: 69 Row: n Column: 19
This clone was selected for full length sequencing because it
passed the following selection criteria: matched mRNA gi: 14249445.
FEATURES
source
1..3452
Location/Qualifiers
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="MGC:44889 IMAGE:5574637"
/tissue_type="Duodenum, adenocarcinoma"
/clone_lib="NIH MGC_88"
/lab_host="DH10B"
/notes="vector: pCMV-SPORT6"
1..3452
/gene="FLJ14431"
/db_xref="LocusID:84869"
159..872
CDS
gene

```

```

/codon_start=1
/product="FLJ14431 protein"
/protein_id="AAH33650.1"
/db_xref="GI:21707067"
/db_xref="LocusID:84869"
/translation="MDKVCAGGGSGIGRAVLAQLMARQYRLAVIARNLGAKAAG
DUGGHLAFSCDAKADHDVQNTFELEKHLGRVNFVLAAGINRDGLVKTEDMVS
QLHTNLGSMLTRKAMRTMIQQQSGIVNVSIGLKGNSQGSVYSKSGGLVGSFR
ALAEVARKIRVNVVAPGVFHTMTKDLKEHLKKNIPLGRFGETIEVAHAVFELLE
SPYITGHVLVDVGLQLIL"
misc_feature
159..857
/note="adh_short; Region: short chain dehydrogenase. This
family contains a wide variety of dehydrogenases"
/db_xref="CDD:pfam00106"

ORIGIN
Alignment Scores:
Pred. No.: 6.82e-10 Length: 3452
Score: 19.00 Matches: 19
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 95.00% Indels: 0
DB: 9 Gaps: 0

US-10-799-747-116 (1-20) x BC033650 (1-3452)

QY 1 MetAlaAlaHisSerValLeuSerPheLeuLeuThrProTyrAlaLeuLysSer 19
|||||
Db 2542 ATGGCAGCCATTTCAGTCTTGAGTTTCTTCTCGACACCTTATGCTCTGAAATCA 2598

RESULT 3
HSM804706
LOCUS Homo sapiens mRNA; cDNA DKFp762K109 (from clone DKFp762K109).
ACCESSION AL833393
VERSION AL833393.1 GI:21734029
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM
REFERENCE
AUTHORS Kyung, K. and Abbott, A.
TITLE The sequence of Homo sapiens BAC clone RP11-483A20
JOURNAL Unpublished (2001)
AUTHORS Waterston, R.H.
TITLE 3 (bases 1 to 175081)
JOURNAL Direct Submission
AUTHORS Submitted (14-JAN-2000) Genome Sequencing Center, Washington
TITLE University School of Medicine, 4444 Forest Park Parkway, St. Louis,
JOURNAL MO 63106, USA
REFERENCE
AUTHORS Waterston, R.H.
TITLE 4 (bases 1 to 175081)
JOURNAL Direct Submission
AUTHORS Submitted (09-AUG-2001) Genome Sequencing Center, Washington
TITLE University School of Medicine, 4444 Forest Park Parkway, St. Louis,
JOURNAL MO 63106, USA
REFERENCE
AUTHORS Waterston, R.
TITLE 5 (bases 1 to 175081)
JOURNAL Direct Submission
AUTHORS Submitted (09-JAN-2002) Department of Genetics, Washington
TITLE University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
JOURNAL On Aug 9, 2001 this sequence version replaced gi:13877272.
COMMENT ----- Genome Center
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc
Contact: sapiens@wustl.edu
----- Summary Statistics
-----
Center project name: H_NH0483A20
-----

NOTICE: This sequence may not represent the entire insert of this
clone. It may be shorter because we only sequence overlapping
clone sections once, or longer because we provide a small overlap
between neighboring data submissions.

This sequence was finished as follows unless otherwise noted:
all regions were double stranded, sequenced with an alternate
chemistry, or covered by high quality data (i.e., phred quality >=
30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by sequence
from more than one subclone; and the assembly was confirmed by
restriction digest.

MAPPING INFORMATION:
Mapping information for this clone was provided by Dr. John D.

```

McPherson, Department of Genetics, Washington University, St. Louis MO. For additional information about the map position of this sequence, see <http://genome.wustl.edu/gsc>

SOURCE INFORMATION:

The RPCI-11 human BAC library was made from the blood of one male donor, as described by Osogawa, K., Woon, P.Y., Zhao, B., Frengen, E., Tateno, M., Catanese, J.J., and de Jong, P.J. (1998) An improved approach for construction of bacterial artificial chromosome libraries. *Genomics* 51:1-8. The clone may be obtained either from Research Genetics, Inc. (<http://www.resgen.com>) or Pieter de Jong and coworkers at the Roswell Park Cancer Institute (<http://bacpac.med.buffalo.edu>)

VECTOR: pBACe3.6

NEIGHBORING SEQUENCE INFORMATION:

The clone sequenced to the right is RP11-36G9. Actual start of this clone is at base position 1 of RP11-483A20; actual end is at base position 175081 of RP11-483A20.

FEATURES

source	Location/Qualifiers
1..175081	
/organism="Homo sapiens"	
/mol_type="Genomic DNA"	
/db_xref="taxon:9606"	
/chromosome="4"	
/map="4"	
/clone="RP11-483A20"	
/clone_lib="RPCI-11"	
1..3267	
/rpt_family="L1"	
3270..3453	
/rpt_family="Alu"	
3434..3464	
/rpt_family="(A)n"	
3455..4161	
/rpt_family="L1"	
7949..7956	
/note="match to EST AW006057 (NID:G5854835) wz81d04.x1"	
10197..10489	
/rpt_family="L1"	
10480..10501	
/rpt_family="AT-rich"	
10495..10665	
/rpt_family="L1"	
10731..11051	
/rpt_family="L1"	
11852..11576	
/rpt_family="L1"	
11564..11598	
/rpt_family="(A)n"	
11642..12324	
/rpt_family="L1"	
11769..11794	
/rpt_family="AT-rich"	
11939..11962	
/rpt_family="AT-rich"	
12335..12412	
/rpt_family="L1"	
12392..12621	
/rpt_family="L1"	
12788..13096	
/rpt_family="Alu"	
13069..13102	
/rpt_family="(CAAA)n"	
13325..13667	
/rpt_family="Alu"	
13444..13484	
/rpt_family="(TAAA)n"	
13640..13686	
/rpt_family="(GAAA)n"	
14249..14422	
/rpt_family="MERI_type"	
14980..15182	
/rpt_family="L1"	
repeat_region	15260..15482
repeat_region	/rpt_family="L1"
repeat_region	15483..15585
repeat_region	/rpt_family="MERI_type"
repeat_region	15986..16105
repeat_region	/rpt_family="L1"
repeat_region	16117..16251
repeat_region	/rpt_family="L1"
repeat_region	16288..16356
repeat_region	/rpt_family="L1"
repeat_region	16373..16545
repeat_region	/rpt_family="L1"
repeat_region	16550..18773
repeat_region	/rpt_family="L1"
repeat_region	18770..20218
repeat_region	/rpt_family="L1"
repeat_region	19228..19346
repeat_region	/rpt_family="T-rich"
repeat_region	20219..20514
repeat_region	/rpt_family="Alu"
repeat_region	20515..22484
repeat_region	/rpt_family="L1"
repeat_region	22584..22882
repeat_region	/rpt_family="Alu"
repeat_region	22748..22768
repeat_region	/rpt_family="AT-rich"
repeat_region	22883..23460
repeat_region	/rpt_family="L1"
repeat_region	23461..23747
repeat_region	/rpt_family="MaLR"
repeat_region	23749..24514
repeat_region	/rpt_family="L1"
misc_feature	24456
misc_feature	/note="match to EST AW902102 (NID:G8066307)"
repeat_region	24488..24797
repeat_region	/note="match to EST AW902102 (NID:G8066307)"
repeat_region	24576..24625
repeat_region	/rpt_family="Mariner"
repeat_region	25484..25787
repeat_region	/rpt_family="Alu"
repeat_region	25758..25787
repeat_region	/rpt_family="(A)n"
repeat_region	25796..26014
repeat_region	/rpt_family="MIR"
repeat_region	26557..26889
repeat_region	/rpt_family="L1"
repeat_region	26890..27205
repeat_region	/rpt_family="MaLR"
repeat_region	27206..27439
repeat_region	/rpt_family="Alu"
repeat_region	27440..27490
repeat_region	/rpt_family="MaLR"
repeat_region	27491..28717
repeat_region	/rpt_family="L1"
repeat_region	28718..28742
repeat_region	/rpt_family="(T)n"
misc_feature	28735..28771
misc_feature	/note="match to EST AA954400 (NID:G3118095) 0001f01.s1"
repeat_region	28747..28826
repeat_region	/note="match to EST AW006057 (NID:G5854835) wz81d04.x1"
repeat_region	29022..29308
repeat_region	/rpt_family="L1"
misc_feature	29384..30105
misc_feature	/note="match to EST BF680971 (NID:G11954866)"
misc_feature	29687..30351
misc_feature	/note="match to EST BG429938 (NID:G1336444)"

Alignment Scores:

Pred. No.:	2,42e-08	Length:	175081
Score:	19.00	Matches:	19
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	95.00%	Indels:	0

DB: 9 Gaps: 0

US-10-799-747-116 (1-20) x AC021151 (1-175081)

QY 1 MetAlaLaHisSerValLeuSerPheLeuLeuTrpThrProTyrAlaLeuLysSer 19
|||||

Db 57641 ATGCGACGCCATTCAGTCTTGAGTTTCTCTCGGACACCTTATGCTCTGAAATCA 57585

RESULT 5
AC117374 281662 bp DNA linear HTG 02-MAY-2002
LOCUS
DEFINITION Homo sapiens chromosome 12 clone RP11-19D19, *** SEQUENCING IN PROGRESS ***, 48 unordered pieces.
AC117374 AC011694
AC117374.1 GI:20127352
KEYWORDS HTG; HTGS PHASE1.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 281662)
Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-Osman,F.R., Allen,C., Alsbrooks,S.L., Amaratunga,H.C., Are,J.R., Ayele,M., Banks,T., Barbacia,J., Benton,J., Bimage,K., Blankenburg,K., Bonnin,D., Bouck,J., Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P., Buhay,C., Burch,P., Burkett,C., Burrell,K.B., Byrd,N.C., Carron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C., Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,X.J., Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R., Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K., Harris,C., Harris,K., Hart,M., Haviak,P., Hawes,A., Hernandez,J., Hernandez,O., Hodgson,A., Hughes,M., Holloway,C., Hollins,B., Homsli,F., Howard,S., Huber,J., Hulyk,S., Hume,J., Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C., Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L., Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,B., Liu,W., Louisleg,H., Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J., Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E., Massey,E., Mawhiney,E., McLeod,M.P., Meador,M., Mei,G., Metzker,M., Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S., Moser,M., Neal,D., Newton,J., Newton,N., Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nwokenwo,S., Oguh,M., Okwuonu,G., Oragunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L., Peters,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y., Rives,M., Rojas,A., Rojubokan,I., Rolfe,M., Ruiz,S., Savery,G., Scherer,S., Scott,G., Shen,H., Shooshtari,N., Sisson,I., Sodergren,E., Sonaik,T., Sparks,A., Stanley,H., Stone,H., Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H., Tansley,J., Taylor,C., Taylor,T., Telford,B., Thomas,N., Thomas,S., Usmani,K., Vasquez,L., Vera,V., Villaford,D., Vinson,R., Wang,Q., Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S., Williams,G., Williamson,A., Wleczyk,R., Wooden,S., Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D., Weinstock,G. and Gibbs,R.

Direct Submission
Unpublished
2 (bases 1 to 281662)
Worley,K.C.
Direct Submission
Submitted (10-APR-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 281662)
Worley,K.C.
Direct Submission
Submitted (02-MAY-2002) Human Genome Sequencing Center, Department

of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
On Apr 28, 2002 this sequence version replaced gi:10045385.

Center: Baylor College of Medicine
Center code: BCM
Web site: <http://www.hgsc.bcm.tmc.edu/>
Contact: hgsc-help@bcm.tmc.edu
Project Information
Center project name: HCN2
Center clone name: RP11-19D19
Summary Statistics
Sequencing vector: Plasmid;
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.990329
Consensus quality: 273063 bases at least Q40
Consensus quality: 294336 bases at least Q30
Consensus quality: 306942 bases at least Q20

* NOTE: Estimated insert size may differ from sequence length
(see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently consists of 48 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

1 2464: contig of 2464 bp in length
2465: gap of unknown length
2565: contig of 2039 bp in length
4603: gap of unknown length
4604: contig of 2410 bp in length
7113: gap of unknown length
7213: gap of unknown length
9517: contig of 2304 bp in length
9518: gap of unknown length
9618: contig of 2490 bp in length
12107: gap of unknown length
12207: contig of 2807 bp in length
15014: gap of unknown length
15114: contig of 2717 bp in length
17831: gap of unknown length
17931: gap of unknown length
20480: contig of 2549 bp in length
20580: gap of unknown length
22700: contig of 2120 bp in length
22800: gap of unknown length
24909: contig of 2109 bp in length
25009: gap of unknown length
27313: contig of 2304 bp in length
27414: gap of unknown length
29798: contig of 2385 bp in length
29898: gap of unknown length
33467: contig of 3568 bp in length
33567: gap of unknown length
36779: contig of 3113 bp in length
36780: gap of unknown length
40472: contig of 3693 bp in length
40573: gap of unknown length
42683: contig of 2111 bp in length
42784: gap of unknown length
46319: contig of 3536 bp in length
46320: gap of unknown length
49507: contig of 3088 bp in length
49607: gap of unknown length
53212: contig of 3605 bp in length
53312: gap of unknown length
57267: contig of 3955 bp in length
57367: gap of unknown length
60444: contig of 3077 bp in length
60445: gap of unknown length
65859: contig of 5315 bp in length
65959: gap of unknown length

```
* 65960 65038: contig of 3079 bp in length
* 69039 65138: gap of unknown length
* 69139 73445: contig of 4307 bp in length
* 73446 73545: gap of unknown length
* 78107 78107: contig of 4562 bp in length
* 78208 78207: gap of unknown length
* 82504 82504: contig of 4297 bp in length
* 82604 82604: gap of unknown length
* 85976 85976: contig of 4372 bp in length
* 87077 87076: gap of unknown length
* 91315 91315: contig of 4239 bp in length
* 91415 91415: gap of unknown length
* 95084 95084: contig of 3669 bp in length
* 95184 95184: gap of unknown length
* 102348 102348: contig of 7164 bp in length
* 102448 102448: gap of unknown length
* 105973 105973: contig of 3525 bp in length
* 105974 106073: gap of unknown length
* 106074 112891: contig of 6818 bp in length
* 112892 112891: gap of unknown length
* 112932 112221: contig of 3230 bp in length
* 116222 116321: gap of unknown length
* 116322 113798: contig of 3477 bp in length
* 119799 119898: gap of unknown length
* 119899 124870: contig of 4972 bp in length
* 124871 124970: gap of unknown length
* 124971 130790: contig of 5820 bp in length
* 130791 130890: gap of unknown length
* 130891 134639: contig of 3749 bp in length
* 134640 134739: gap of unknown length
* 134740 140637: contig of 5898 bp in length
* 140638 140737: gap of unknown length
* 140738 145541: contig of 4804 bp in length
* 145542 145641: gap of unknown length
* 145642 155057: contig of 9416 bp in length
* 155058 155157: gap of unknown length
* 155158 164357: contig of 9200 bp in length
* 164358 164457: gap of unknown length
* 164458 174892: contig of 10435 bp in length
* 174893 174992: gap of unknown length
* 174993 184400: contig of 9408 bp in length
* 184401 184500: gap of unknown length
* 184501 198385: contig of 13885 bp in length
* 198386 198485: gap of unknown length
* 198486 216436: contig of 17951 bp in length
* 216437 216536: gap of unknown length
* 22916 232916: contig of 16380 bp in length
* 232917 233016: gap of unknown length
* 233017 255437: contig of 22421 bp in length
* 255438 255537: gap of unknown length
* 255538 281662: contig of 26125 bp in length.
```

FEATURES

```
source
1. .281662
Location/Qualifiers
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="12"
/clone="RP11-19D19"
```

ORIGIN

```
Alignment Scores:
Pred. No.: 3.73e-08 Length: 281662
Score: 19.00 Matches: 19
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 95.00% Indels: 0
DB: 2 Gaps: 0
```

US-10-799-747-116 (1-20) x AC117374 (1-281662)

Qy 1 MetAlaLaHisSerValLeuSerPheIeuLeuTrpThrProTyAlaLeuIysSer 19

Db 145718 ATGGCAGCCCATTCAGTCTTCTGAGTTTCTTCTCTGGACCATATGCTCTGAAATCA 145662

```
RESULT 6
HSM800227 2020 bp mRNA linear PRI 18-FEB-2000
LOCUS Homo sapiens mRNA; cDNA DKFp586N1720 (from clone DKFp586N1720).
DEFINITION
ACCESSION AL049442
VERSION AL049442.1 GI:4500222
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 2020)
AUTHORS Wambutt,R., Heubner,D., Mewes,H.W., Gassenhuber,J. and Wiemann,S.
TITLE Direct Submission
JOURNAL Submitted (10-MAR-1999) MIPS, Am Klopferspitz 18a, D-82152
Martinsried, GERMANY
COMMENT Clone from S. Wiemann, Molecular Genome Analysis, German Cancer
Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de;
sequenced by AGOWA (Berlin/Germany) within the CDNA sequencing
consortium of the German Genome Project.
This clone (DKFp586N1720) is available at the RZPD in Berlin.
Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059
Berlin-Charlottenburg, GERMANY; Email: clone@rzd.de Further
information about the clone and the sequencing project is available
at http://www.mips.biochem.mpg.de/proj/cDNA/.
```

FEATURES

```
source
1. .2020
Location/Qualifiers
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="RZPD:DKFp586N1720"
/db_xref="taxon:9606"
/clone="DKFp586N1720"
/tissue_type="uterus"
/clone_lib="586 (synonym: hutel). Vector pSport1; host
DH10B; sites NotI + SalI/MluI"
/dev stage="adult"
polyA_signal 1983..1988
polyA_site 2003
```

ORIGIN

```
Alignment Scores:
Pred. No.: 0.000538 Length: 2020
Score: 13.00 Matches: 13
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 65.00% Indels: 0
DB: 9 Gaps: 0
```

US-10-799-747-116 (1-20) x HSM800227 (1-2020)

Qy 1 MetAlaLaHisSerValLeuSerPheIeuLeuTrpThr 13

Db 1115 ATGGCAGCCCATTCAGTCTTCTGAGTTTCTTCTCTGGACA 1153

RESULT 7

```
AC123991/c
LOCUS AC123991 153464 bp DNA linear HTG 17-MAR-2003
DEFINITION Mus musculus clone RP24-137I10, WORKING DRAFT SEQUENCE, 3 unordered
pieces.
```

```
ACCESSION AC123991
VERSION AC123991.3 GI:28975888
KEYWORDS HTG; HTGS PHASE1; HTGS DRAFT.
SOURCE Mus musculus (house mouse)
```

```
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE
1 (bases 1 to 153464)
AUTHORS Birren,B., Nusbaum,C. and Lander,E.
```

TITLE Mus musculus, clone RP24-137I10

JOURNAL

Unpublished

REFERENCE

2 (bases 1 to 153464)

AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,

Anderson, S., Barna, N., Bastien, V., Bloom, T., Boguslavskiy, L.,
 Boukhgalter, B., Brown, A., Camarata, J., Campopiano, A., Chang, J.,
 Chazaro, B., Choepel, Y., Collangelo, M., Collins, S., Collymore, A.,
 Cook, A., Cooke, P., DeArellano, K., Dewar, K., Diaz, J.S., Dodge, S.,
 Faro, S., Ferreira, P., Fitzgerald, M., FitzHugh, W., Gage, D.,
 Galagan, J., Gardyna, S., Ginde, S., Gord, S., Goyette, M., Graham, L.,
 Grand-pierre, N., Hagos, B., Horton, L., Hulme, W., Iliev, I.,
 Johnson, R., Jones, C., Kama, A., Karatas, A., Kells, C., LaRoque, K.,
 Lamazares, R., Landers, T., Lehoczy, J., Levine, R., Lindblad-Toh, K.,
 Liu, G., MacLean, C., MacDonald, P., Major, J., Marquis, N.,
 Matthews, C., McCarthy, M., McEwan, P., McKernan, K., Meldrim, J.,
 Meneus, L., Mihova, T., Menga, V., Murphy, T., Naylor, J., Nguyen, C.,
 Nicol, R., Norbu, C., Norman, C.H., O'Connor, T., O'Donnell, P.,
 O'Neill, D., Oliver, J., Peterson, K., Phunkhang, P., Pierre, N.,
 Pollara, V., Raymond, C., Retta, R., Rieback, M., Riley, R., Rise, C.,
 Rogov, P., Roman, J., Rosetti, M., Roy, A., Santos, R., Schauer, S.,
 Schupback, R., Seaman, S., Severy, P., Spencer, B., Stange-Thomann, N.,
 Stojanovic, N., Strauss, N., Subramanian, A., Talamas, J., Tesfaye, S.,
 Theodore, J., Topham, K., Travers, M., Travis, N., Trigilio, J.,
 Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J.,
 Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

Direct Submission
 Submitted (06-JUN-2002) Whitehead Institute/MIT Center for Genome
 Research, 320 Charles Street, Cambridge, MA 02141, USA
 3 (bases 1 to 153464)

Birren, B., Nusbaum, C., Lander, E., Abouelleil, A., Allen, N.,
 Anderson, S., Arachchi, H.M., Barna, N., Bastien, V., Bloom, T.,
 Boguslavskiy, L., Boukhgalter, B., Camarata, J., Chang, J., Choepel, Y.,
 Collymore, A., Cook, A., Cooke, P., Corum, B., DeArellano, K.,
 Diaz, J.S., Dodge, S., Dooley, K., Dorris, L., Erickson, J., Faro, S.,
 Ferreira, P., Fitzgerald, M., Gage, D., Galagan, J., Gardyna, S.,
 Graham, L., Grand-Pierre, N., Hafez, N., Hagopian, D., Hagos, B.,
 Hall, L., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C.,
 Kama, A., Karatas, A., Kells, C., Landers, T., Levine, R.,
 Lindblad-Toh, K., Liu, G., Lui, A., Mabbitt, R., MacLean, C.,
 MacDonald, P., Major, J., Manning, J., Matthews, C., McCarthy, M.,
 Meldrim, J., Meneus, L., Mihova, T., Menga, V., Murphy, T., Naylor, J.,
 Nguyen, C., Nicol, R., Norbu, C., O'Connor, T., O'Donnell, P.,
 O'Neill, D., Oliver, J., Peterson, K., Phunkhang, P., Pierre, N.,
 Rachupka, A., Ramasamy, U., Raymond, C., Retta, R., Rise, C., Rogov, P.,
 Roman, J., Schauer, S., Schupback, R., Seaman, S., Severy, P., Smith, C.,
 Spencer, B., Stange-Thomann, N., Stojanovic, N., Stubbs, M.,
 Talamas, J., Tesfaye, S., Theodore, J., Topham, K., Travers, M.,
 Vassiliev, H., Venkataraman, V.S., Viel, R., Vo, A., Wilson, B., Wu, X.,
 Wyman, D., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

Direct Submission
 Submitted (17-MAR-2003) Whitehead Institute/MIT Center for Genome
 Research, 320 Charles Street, Cambridge, MA 02141, USA
 On Mar 17, 2003 this sequence version replaced gi:28201678.
 All repeats were identified using RepeatMasker:
 Smit, A.F.A. & Green, P. (1996-1997)
 http://ftp.genome.washington.edu/RM/RepeatMasker.html

----- Genome Center
 Center: Whitehead Institute/ MIT Center for Genome Research
 Center code: WITBR
 Web site: <http://www-seq.wi.mit.edu>
 Contact: sequence_submissions@genome.wi.mit.edu

----- Project Information
 Center project name: L26544
 Center clone name: 137 I.10
 ----- Summary Statistics

Sequencing vector: Plasmid; n/a; 100% of reads
 Chemistry: Dye-terminator Big Dye; 100% of reads
 Assembly program: Phrap; version 0.960731
 Consensus quality: 152581 bases at least Q40
 Consensus quality: 152929 bases at least Q30
 Consensus quality: 153074 bases at least Q20
 Insert size: 151000; agarose-fp
 Insert size: 153284; sum-of-ctontigs
 Quality coverage: 11.8 in Q20 bases; agarose-fp
 Quality coverage: 11.7 in Q20 bases; sum-of-ctontigs

 * NOTE: This is a 'working draft' sequence. It currently

* consists of 3 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

* 1 132252: contig of 132252 bp in length
 * 132253 132352: gap of 100 bp
 * 132353 134109: contig of 1757 bp in length
 * 134110 134209: gap of 100 bp
 * 134210 153464: contig of 19255 bp in length.

FEATURES

source

1. 153464
 /organism="Mus musculus"
 /mol_type="genomic DNA"
 /db_xref="taxon:10090"
 /clone="RP24-137110"
 /clone_lib="RPCI-24 Male Mouse BAC"
 1. 132252
 /note="assembly_fragment
 clone_end:SP6
 vector_side:left"
 132353. 134109
 /note="assembly_fragment"
 134210. 153464
 /note="assembly_fragment
 clone_end:T7
 vector_side:right"

misc_feature

1. 132252
 /note="assembly_fragment
 clone_end:SP6
 vector_side:left"

132353. 134109

/note="assembly_fragment"

134210. 153464

/note="assembly_fragment
 clone_end:T7
 vector_side:right"

134210. 153464

/note="assembly_fragment
 clone_end:T7
 vector_side:right"

134210. 153464

/note="assembly_fragment
 clone_end:T7
 vector_side:right"

134210. 153464

/note="assembly_fragment
 clone_end:T7
 vector_side:right"

134210. 153464

/note="assembly_fragment
 clone_end:T7
 vector_side:right"

134210. 153464

/note="assembly_fragment
 clone_end:T7
 vector_side:right"

134210. 153464

/note="assembly_fragment
 clone_end:T7
 vector_side:right"

134210. 153464

/note="assembly_fragment
 clone_end:T7
 vector_side:right"

134210. 153464

/note="assembly_fragment
 clone_end:T7
 vector_side:right"

134210. 153464

/note="assembly_fragment
 clone_end:T7
 vector_side:right"

134210. 153464

/note="assembly_fragment
 clone_end:T7
 vector_side:right"

134210. 153464

/note="assembly_fragment
 clone_end:T7
 vector_side:right"

134210. 153464

/note="assembly_fragment
 clone_end:T7
 vector_side:right"

134210. 153464

/note="assembly_fragment
 clone_end:T7
 vector_side:right"

134210. 153464

/note="assembly_fragment
 clone_end:T7
 vector_side:right"

134210. 153464

/note="assembly_fragment
 clone_end:T7
 vector_side:right"

134210. 153464

/note="assembly_fragment
 clone_end:T7
 vector_side:right"

134210. 153464

/note="assembly_fragment
 clone_end:T7
 vector_side:right"

134210. 153464

/note="assembly_fragment
 clone_end:T7
 vector_side:right"

134210. 153464

/note="assembly_fragment
 clone_end:T7
 vector_side:right"

134210. 153464

/note="assembly_fragment
 clone_end:T7
 vector_side:right"

134210. 153464

/note="assembly_fragment
 clone_end:T7
 vector_side:right"

134210. 153464

/note="assembly_fragment
 clone_end:T7
 vector_side:right"

134210. 153464

/note="assembly_fragment
 clone_end:T7
 vector_side:right"

134210. 153464

/note="assembly_fragment
 clone_end:T7
 vector_side:right"

EM, EMBL, SW, SWISSPROT, Tr., TREMBL; Wp., WORMPEP; Information on the WORMPEP database can be found at http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence was generated from part of bacterial clone contigs of human chromosome 20, constructed by the Sanger Centre Chromosome 20 Mapping Group. Further information can be found at <http://www.sanger.ac.uk/HGP/Chr20>

This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest. RP5-828H9 is from the library RPCI-5 constructed by the group of Pieter de Jong. For further details see <http://www.chori.org/bacpac/home.htm>

VECTOR: pCYPAC2

IMPORTANT: This sequence is not the entire insert of clone RP5-828H9 It may be shorter because we sequence overlapping sections only once, except for a 100 base overlap.

The left end of clone RP4-680N4 is at 166918 in this sequence. The true right end of clone RP5-1022P6 is at 99 in this sequence. The true right end of clone RP4-766D4 is at 46185 in this sequence.

FEATURES

source

```

1..167013
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="20"
/clone="RP5-828H9"
/clone_lib="RPCI-5"
785..1214
/feature="HERV17 repeat: matches 6052..6478 of consensus"
1271..1441
/feature="match: GSS: Em:AQ780632"
1336..1566
/feature="match: GSS: Em:AQ780632"
/feature="MER66-internal repeat: matches 4624..4847 of consensus"
1446..1607
/feature="match: GSS: Em:AQ780632"
1798..1923
/feature="MER51-internal repeat: matches 7690..7816 of consensus"
2182..2263
/feature="MER51B repeat: matches 190..523 of consensus"
2375..2451
/feature="AluS repeat: matches 1..67 of consensus"
2455..2627
/feature="AluM4 repeat: matches 5635..5810 of consensus"
2628..2935
/feature="AluSq repeat: matches 1..308 of consensus"
2936..3001
/feature="AluM4 repeat: matches 5569..5635 of consensus"
3002..3305
/feature="AluJb repeat: matches 1..300 of consensus"
3306..3764
/feature="AluM4 repeat: matches 5120..5569 of consensus"
3765..4077
/feature="AluJo repeat: matches 1..292 of consensus"
4078..4223
/feature="AluM4 repeat: matches 4973..5120 of consensus"
4224..4361
/feature="AluJo/FPAM repeat: matches 158..294 of consensus"
/feature="complement(join(4666..4827,5364..5428,7817..7945))"
/feature="complement(join(4666..4827,5364..5428,7817..7945))"
/feature="product="dJ828H9.2 (putative novel transcript)"
/feature="match: ESTs: Em:A1218423"
/feature="evidence=not experimental"
/feature="complement(4666)"
/feature="dJ828H9.2"

```

```

repeat_region
/feature="AluSx repeat: matches 1..298 of consensus"
5519..5794
/feature="L2 repeat: matches 2405..2680 of consensus"
5934..6244
/feature="match: GSS: Em:AQ415067"
6021..6321
/feature="match: GSS: Em:AQ531024"
6258..7406
/feature="L1P4S repeat: matches 4995..6144 of consensus"
8230..8343
/feature="MIR repeat: matches 45..166 of consensus"
/feature="complement(join(9193..9766,11907..12175))"
/feature="dJ828H9.1"
/feature="complement(join(9193..9766,11907..12175))"
/feature="dJ828H9.1"
/feature="product="dJ828H9.1 (putative novel transcript)"
/feature="match: ESTs: Em:AA431369 Em:AA382348 Em:AA397834 Em:AA758864 Em:AA906363 Em:AA757279 Em:AW341461 Em:AA432366"
/feature="evidence=not experimental"
/feature="complement(9193)"
/feature="dJ828H9.1"
/feature="complement(9426)"
/feature="dJ828H9.1"
10546..10649
/feature="L2 repeat: matches 541..654 of consensus"
10708..10784
/feature="ORSL repeat: matches 206..275 of consensus"
10785..11076
/feature="AluY repeat: matches 2..295 of consensus"
11077..11230
/feature="ORSL repeat: matches 275..458 of consensus"
/feature="complement(12059..12555)"
13524..13668
/feature="MER5B repeat: matches 17..166 of consensus"
14312..14516
/feature="AluJb repeat: matches 111..306 of consensus"
14691..15441
/feature="L1MB3 repeat: matches 5264..6016 of consensus"
15414..15494
/feature="L1ME3A repeat: matches 6025..6104 of consensus"
15495..15805
/feature="AluSq repeat: matches 1..309 of consensus"
15976..16160
/feature="MIRIG repeat: matches 39..229 of consensus"
16253..16343
/feature="MER5A repeat: matches 58..145 of consensus"
16275..16389
/feature="MER5A repeat: matches 1..123 of consensus"
16406..16912
/feature="match: GSS: Em:AQ792876"
16434..16597
/feature="MIR repeat: matches 91..262 of consensus"
17457..17911
/feature="match: GSS: Em:AQ135214"
17487..17994
/feature="match: GSS: Em:AQ215088"
17672..18210
/feature="MER52A repeat: matches 1207..1755 of consensus"
18311..18921
/feature="MER52A repeat: matches 660..1349 of consensus"
18976..19568
/feature="MER4-internal repeat: matches 4586..5102 of consensus"
19578..24868
/feature="L1P4S repeat: matches 10..5050 of consensus"
24869..25168
/feature="AluSq repeat: matches 1..300 of consensus"
25169..26305
/feature="L1P4S repeat: matches 5050..6163 of consensus"
26308..26385

```



```

/note="MER4-internal repeat: matches 3313. .3393 of
consensus"
repeat_region 26386. .26686
/note="AluJb repeat: matches 1. .302 of consensus"
repeat_region 26687. .26832
/note="MER4-internal repeat: matches 3393. .3543 of
consensus"
repeat_region 26833. .27129
/note="AluXx repeat: matches 1. .296 of consensus"
repeat_region 27130. .27559
/note="MER4-internal repeat: matches 3543. .3989 of
consensus"
repeat_region 27560. .27865
/note="AluJb repeat: matches 1. .301 of consensus"
repeat_region 27866. .28012
/note="MER4-internal repeat: matches 3989. .4140 of
consensus"
repeat_region 28013. .28262
/note="AluJc repeat: matches 3. .264 of consensus"
repeat_region 28263. .28514
/note="MER4-internal repeat: matches 4140. .4401 of
consensus"
repeat_region 28515. .28797
/note="AluXx repeat: matches 1. .284 of consensus"
repeat_region 28798. .29064
/note="MER4-internal repeat: matches 4401. .6347 of
consensus"
misc_feature 28993. .29239

```

Alignment Scores:

```

Pred. No.: 353 Length: 167013
Score: 9.00 Matches: 9
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 45.00% Indels: 0
DB: 9 Gaps: 0

```

US-10-799-747-116 (1-20) x HSDJ828H9 (1-167013)

Qy 5 ServalleuSerPheLeuLeuTyrThr 13

Db 84047 TCTGTGCTCAGCTTCTCTCTGGACA 84021

RESULT 9

AC122809/c
LOCUS AC122809 205725 bp DNA linear ROD 25-NOV-2003
DEFINITION Mus musculus BAC clone RP23-272E5 from chromosome 7, complete
sequence.

ACCESSION AC122809

VERSION AC122809.4 GI:33300789

KEYWORDS HTG.

SOURCE Mus musculus (house mouse)

ORGANISM Mus musculus

REFERENCE 1 (bases 1 to 205725)
Bukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.

AUTHORS Sweatengen-Shanid, S. and Meyer, R.

TITLE The sequence of Mus musculus BAC clone RP23-272E5

JOURNAL Unpublished (2001)

REFERENCE 2 (bases 1 to 205725)

AUTHORS Wilson, R.

TITLE Sequencing of Mus musculus

JOURNAL Unpublished (2001)

REFERENCE 3 (bases 1 to 205725)

AUTHORS McPherson, J.D. and Waterston, R.H.

TITLE Direct Submission

JOURNAL Submitted (25-MAY-2002) Genome Sequencing Center, 4444 Forest Park

REFERENCE 4 (bases 1 to 205725)

AUTHORS McPherson, J.D. and Waterston, R.H.

TITLE Direct Submission

JOURNAL Submitted (13-FEB-2003) Genome Sequencing Center, 4444 Forest Park

REFERENCE 5 (bases 1 to 205725)

REFERENCE
AUTHORS
TITLE
JOURNAL

5 (bases 1 to 205725)

Wilson, R.K.

Direct Submission

Submitted (29-JUL-2003) Genome Sequencing Center, 4444 Forest Park
Parkway, St. Louis, MO 63108, USA

6 (bases 1 to 205725)

Wilson, R.

Direct Submission

Submitted (25-NOV-2003) Department of Genetics, Washington
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
On Jul 29, 2003 this sequence version replaced gi:28372763.

----- Genome Center

Center: Washington University Genome Sequencing Center

Center code: WUGSC

Web site: <http://genome.wustl.edu>

Contact: submissions@watson.wustl.edu

----- Summary Statistics

----- Center project name: M_BA0272E05

NOTICE: This sequence may not represent the entire insert of this
clone. It may be shorter because we only sequence overlapping
clone sections once, or longer because we provide a small overlap
between neighboring data submissions.

This sequence was finished as follows unless otherwise noted:
all regions were double stranded, sequenced with an alternate
chemistry, or covered by high quality data (i.e., phred quality >=
30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by sequence
from more than one subclone; and the assembly was confirmed by
restriction digest.

MAPPING INFORMATION:

Mapping information for this clone was provided by Dr. Wes Warren,
Department of Genetics, Washington University, St. Louis MO. For
additional information about the map position of this sequence, see
<http://genome.wustl.edu>

SOURCE INFORMATION:

The RPECI-23 BAC Library has been constructed by Kazutoyo Osegawa
and Minako Tateno in the laboratory of Pieter de Jong
(<http://www.chori.org>) from female C57BL/6J mouse kidney and/or
brain genomic DNA. The clone and detailed information can be
obtained from Research Genetics, Inc. (<http://www.resgen.com>) or
Pieter de Jong and coworkers at <http://www.chori.org>

NEIGHBORING SEQUENCE INFORMATION:

This sequence is the entire insert of the clone.

FEATURES
source

Location/Qualifiers

1. .205725

/organism="Mus musculus"

/mol_type="genomic DNA"

/db_xref="taxon:10090"

/chromosome="7"

/map="7"

/clone="RP23-272E5"

/clone_lib="RPECI-23"

1. .94

/rpt_family="Alu"

92. .106

/rpt_family="MaLR"

107. .295

/rpt_family="B2"

296. .525

/rpt_family="MaLR"

706. .778

/rpt_family="Alu"

831. .960

/rpt_family="MIR"

1158. .1200

/rpt_family="Alu"

1201. .1596

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region	/rpt_family="MaLR" 1597.1667
repeat_region	/rpt_family="Alu" 1817.1930
repeat_region	/rpt_family="Alu" 2015.2069
repeat_region	/rpt_family="ID" 2694.2809
repeat_region	/rpt_family="Alu" 2866.3093
repeat_region	/rpt_family="B2" 3155.3231
repeat_region	/rpt_family="ERV1" 3302.3431
repeat_region	/rpt_family="Alu" 4114.4317
repeat_region	/rpt_family="L1" 4320.4395
repeat_region	/rpt_family="L1" 4396.4542
repeat_region	/rpt_family="Alu" 4543.4609
repeat_region	/rpt_family="L1" 4610.4759
repeat_region	/rpt_family="Alu" 4760.4898
repeat_region	/rpt_family="L1" 4899.5280
repeat_region	/rpt_family="MaLR" 5281.6123
repeat_region	/rpt_family="L1" 6129.6149
repeat_region	/rpt_family="MaLR" 6150.6253
repeat_region	/rpt_family="Alu" 6340.6478
repeat_region	/rpt_family="Alu" 8595.8732
repeat_region	/rpt_family="Alu" 8626.8823
repeat_region	/rpt_family="MaLR" 9217.9562
repeat_region	/rpt_family="MaLR" 15551.15699
repeat_region	/rpt_family="Alu" 15734.15935
repeat_region	/rpt_family="Alu" 16936.17076
repeat_region	/rpt_family="Alu" 18004.18146
repeat_region	/rpt_family="Alu" 18164.18345
repeat_region	/rpt_family="B2" 18848.18921
repeat_region	/rpt_family="ID" 18938.19051
repeat_region	/rpt_family="Alu" 19472.19645
repeat_region	/rpt_family="B2" 19689.19795
repeat_region	/rpt_family="Alu" 20808.20887
repeat_region	/rpt_family="L1" 21309.21329
repeat_region	/rpt_family="B4" 21330.21417
repeat_region	/rpt_family="Alu" 21436.21586
repeat_region	/rpt_family="MaLR" 21557.21757
repeat_region	/rpt_family="MaLR" 21771.21812
repeat_region	/rpt_family="Alu"

repeat_region	21948..22091
rpt_family="Alu"	
repeat_region	22448..23281
rpt_family="L1"	
repeat_region	23288..23405
rpt_family="Alu"	
repeat_region	23447..23515
rpt_family="ID"	
repeat_region	23839..24523
rpt_family="RMR19A"	
repeat_region	24541..24659
rpt_family="L1"	
repeat_region	24690..24764
rpt_family="ID"	
repeat_region	25308..25445
rpt_family="Alu"	
repeat_region	26307..26322
rpt_family="Alu"	
repeat_region	27425..27598
rpt_family="B4"	
repeat_region	27620..27673
rpt_family="ID"	
repeat_region	27674..27706
rpt_family="B4"	
repeat_region	27703..27841
rpt_family="B2"	

Alignment Scores:

Pred. No.:	426	Length:	205725
Score:	9.00	Matches:	9
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	45.00%	Indels:	0
DR:	10	Gaps:	0

US-10-799-747-116 (1-20) x AC122809 (1-205725)

QY 3 AlaHisSerValLeuSerPheLeuLeu 11
|||
pb 136760 GCACATTCACATTTTGTCATCTTGCTA 136734

RESULT 10

Accession	LOCUS	DEFINITION	Linear	HTG 09-NOV-2002
AC110315	231768 bp	Rattus norvegicus clone CH230-191n1, *** SEQUENCING IN PROGRESS	linear	
AC110315		***. 2 unordered pieces.		

ACCESSION	AC110315
VERSION	AC110315.5
KEYWORDS	GI:24818671
KEYWORDS	HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ENRICHED
SOURCE	Rattus norvegicus (Norway rat)
ORGANISM	Rattus norvegicus

ORGANISM

Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus

REFERENCE 1 (bases 1 to 231768)
Kucenas.

REFERENCES

Muzny, D. Marie, Metzker, M. Lee, Abramson, S., Adams, C., Alder, J., Allen, C., Allen, H., Alsbrooks, S., Amin, A., Anguiano, D., Anyalebechi, V., Aoyagi, A., Ayodeji, M., Baca, E., Baden, H., Baldwin, D., Bandaranaike, D., Barber, M., Barnstead, M., Benahmed, F., Biswal, K., Blair, J., Blankenburg, K., Blyth, P., Brown, M., Bryant, N., Buhay, C., Burch, P., Burrell, K., Calderon, E., Cardenas, V., Carter, K., Cavazos, I., Ceasar, H., Center, A., Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Y., Chen, Z., Chu, J., Cleveland, C., Cockrell, R., Cox, C., Coyle, M., Cree, A., D'Souza, L., Davila, M. L., Davis, C., Davy-Carroll, L., De Anda, C., Dederich, D., Delgado, O., Denson, S., Deramo, C., Ding, Y., Dinh, H., Divya, K., Draper, H., Dugan-Rocha, S., Dunn, A., Durbin, K., Duvall, B., Eaves, K., Egan, A., Escotto, M., Eugene, C., Evans, C. A., Falls, T., Fan, G., Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P., Fraser, C. M., Gabisi, A., Ganta, R., Garcia, A., Garner, T., Garza, W., Gebregregorios, E., Geer, K., Gall, R., Grady, M., Guerra, W., Gunaratne, P., Haaland, W., Hamill, C., Hamilton, C., Hamilton, K., Harvey, Y., Havlak, P., Hawes, A., Henderson, N., Hernandez, J.,

Hernandez, R., Hines, S., Hladun, S.L., Hodgson, A., Hogues, M., Hollins, B., Howells, S., Hulyk, S., Hume, J., Idlebird, D., Jackson, A., Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolivet, A., Karpathy, S., Kelly, S., Kelly, S., Khan, Z., King, L., Kovar, C., Kowis, C., Kraft, C.L., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, J., Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J., Lorensuhewa, L., Loulseged, H., Lozado, R.J., Lu, X., Ma, J., Maheshwari, M., Mahindartne, M., Mahmoud, M., Malloy, K., Mangum, A., Mangum, B., Mapua, P., Martin, K., Martin, R., Martinez, E., Mawhinney, S., McLeod, M.P., McNeill, T.Z., Meenen, E., Milosavljevic, A., Miner, G., Minja, E., Montemayor, J., Moore, S., Morgan, M., Morris, K., Morris, S., Munidas, M., Murphy, M., Nair, L., Nankervis, C., Neal, D., Newton, N., Nguyen, N., Norris, S., Parks, K., Nwaokelimeh, O., Okwuonu, G., Olampunsegon, A., Pal, S., Pankoch, C., Pasternak, S., Paul, H., Perez, A., Perez, L., Pfannkuch, C., Plopper, F., Poindexter, A., Popovic, D., Primus, E., Pu, L., Puazo, M., Quiroz, J., Rachlin, E., Reeves, K., Regier, M.A., Reigh, R., Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, F., Rives, C., Rodkey, T., Rojas, A., Rose, M., Rose, R., Ruiz, S.J., Sanders, W., Savary, G., Scherer, S., Scott, G., Shatsman, S., Shen, H., Shetty, J., Shvartsbeyn, A., Sisson, I., Sitter, C.D., Smajs, D., Sneed, A., Sodergren, E., Song, X.-Z., Sorelle, R., Sosa, J., Steimle, M., Strong, R., Sutton, A., Svatek, A., Tabor, P., Taylor, C., Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Usmani, K., Valas, R., Vera, V., Villasana, D., Waldron, L., Walker, B., Wang, J., Wang, Q., Wang, S., Warren, R., Warren, R., Wei, X., White, F., Williams, G., Willson, R., Wleczek, R., Wooden, H., Worley, K., Wright, D., Wright, J., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V., Yu, F., Zhang, J., Zhou, J., Zhou, X., Zhao, S., Zhao, S., Dunn, D., von Niederhausern, A., Weiss, R., Smith, D.R., Holt, R.A., Smith, H.O., Weinstock, G. and Gibbs, R.A.

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

Unpublished
2 (bases 1 to 231768)
Worley, K.C.

Direct Submission

Submitted (11-FEB-2002)

Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

3 (bases 1 to 231768)

Rat Genome Sequencing Consortium.

Direct Submission

Submitted (09-NOV-2002)

Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

On Nov 9, 2002 this sequence version replaced gi:23265946.

The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (<http://www.hgsc.bcm.tmc.edu/projects/rat/>). Each contig described in the feature table below represents a scaffold in the Atlas assembly (a 'contig-scaffold'). Within each contig-scaffold, individual sequence contigs are ordered and oriented, and separated by sized gaps filled with Ns to the estimated size. The sequence may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.

----- Genome Center

Center: Baylor College of Medicine

Center code: BCM

Web site: <http://www.hgsc.bcm.tmc.edu/>

Contact: hgsc-help@bcm.tmc.edu

----- Project Information

Center project name: GOGP

Center clone name: CH230-191N21

----- Summary Statistics

Assembly program: Phrap; version 0.990329

Consensus quality: 218739 bases at least Q40

Consensus quality: 220995 bases at least Q30

Consensus quality: 224005 bases at least Q20

Estimated insert size: 223718; sum-of-contigs estimation

Quality coverage: 6x in Q20 bases; sum-of-contigs estimation

* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 2 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

* 1 230654: contig of 230654 bp in length
* 230655 230754: gap of unknown length
* 230755 231768: contig of 1014 bp in length.

FEATURES

Source

1..231768
/organism="Rattus norvegicus"
/mol_type="genomic DNA"
/db_xref="taxon:10116"
/clone="CH230-191N21"
1..1220
/note="wgs_end_extension"
clone_end:5p6"
1824..2452
/note="clone_boundary"
clone_end:5p6"
site:EcoRI
end sequence:BH282813"
complement(222634..223294)
/note="clone_boundary"
clone_end:T7
site:EcoRI
end sequence:BH282812"
226573..227843
/note="wgs_end_extension"
clone_end:T7"
228778..230654
/note="wgs_end_extension"
clone_end:T7"

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

ORIGIN

Alignment Scores:

Pred. No.:

Score: 475

Percent Similarity: 9.00

Best Local Similarity: 100.00%

Query Match: 100.00%

Indels: 45.00%

DB: 2

Gaps: 0

US-10-799-747-116 (1-20) x AC110315 (1-231768)

QY 6 ValLeuSerPheLeuLeuTriThrPro 14

Db 172304 GTGTTTCCTTCCTGTTATGACTCCA 172330

RESULT 11

AC103176/c

LOCUS

AC103176

DEFINITION

Rattus norvegicus clone CH230-98N18, WORKING DRAFT SEQUENCE, 4

unordered pieces.

AC103176

AC103176.5 GI:30578547

HTG; HTGS PHASE1; HTGS DRAFT; HTGS_FULLTOP.

KEYWORDS

Rattus norvegicus (Norway rat)

SOURCE

ORGANISM

Rattus norvegicus

Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;

Rattus.

1 (bases 1 to 254839)

Muzny, D.Marie., Metzker, M.Lee., Abramzon, S., Adams, C., Alder, J.,

Allen, C., Allen, H., Alsbrooks, S., Amin, A., Anguiano, D.,

Ayalebechi, V., Aoyagi, A., Ayodeji, M., Baca, E., Baden, H.,

Baldwin, D., Bandaranaike, D., Barber, M., Barnstead, M., Benahmed, F.,

AC103176 254839 bp DNA linear HTG 13-MAY-2003

Biewalo, K., Blair, J., Blankenburg, K., Blyth, P., Brown, M., Bryant, N., Buhay, C., Burch, P., Burrell, K., Calderon, E., Cardenas, V., Carter, K., Cavazos, I., Ceasar, H., Center, A., Chacko, J., Chavez, D., Chen, G., Chen, Y., Chen, Y., Chu, J., Cleveland, C., Cockrell, R., Cox, C., Coyle, M., Cree, A., D'Souza, L., Davila, M.L., Davis, C., Davy-Carroll, L., De Anda, C., Dederich, D., Delgado, O., Denson, S., Deramo, C., Ding, Y., Dinh, H., Divya, K., Draper, H., Dugan-Rocha, S., Dunn, A., Durbin, K., Duval, B., Egan, A., Escotto, M., Eugene, C., Evans, C.A., Falls, T., Fan, G., Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P., Fraser, C.M., Gabisi, A., Ganta, R., Garcia, A., Garner, T., Garza, M., Gebregorgis, E., Geer, K., Gill, R., Grady, M., Guerra, W., Guevara, W., Gunaratne, P., Haaland, W., Hamil, C., Hamilton, C., Hamilton, K., Harvey, Y., Havlak, P., Hawes, A., Henderson, N., Hernandez, J., Hernandez, R., Hines, S., Hladun, S.L., Hodgson, A., Hognes, M., Hollins, B., Howells, S., Hulyk, S., Hume, J., Idlebird, D., Jackson, A., Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolivet, A., Karpathy, S., Kelly, S., Kelly, S., Khan, Z., King, L., Kovar, C., Kowis, C., Kraft, C.L., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, J., Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J., Lorensheva, L., Louiseged, H., Lozado, R.J., Lu, X., Ma, J., Maheshwari, M., Mahindartne, M., Mahmoud, M., Malloy, K., Mangum, A., Mangum, B., Mapa, P., Martin, K., Martin, R., Martinez, E., Mathiney, S., McLeod, M.P., McNeill, T.Z., Meenen, E., Milosavljevic, A., Miner, G., Minja, E., Montemayor, J., Moore, S., Morgan, M., Morris, K., Morris, S., Munidasa, M., Murphy, M., Nair, L., Nankervis, C., Neal, D., Newton, N., Nguyen, N., Norris, S., Nwaokemele, O., Okwuonu, G., Olarnpunsagoon, A., Pal, S., Parks, K., Pasternak, S., Paul, H., Perez, A., Perez, L., Pfannkoch, C., Plapper, F., Poindexter, A., Popovic, D., Primus, E., Pu, L.-L., Puafo, M., Quiroz, J., Rachlin, E., Reeves, K., Regier, M.A., Reigh, R., Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, F., Rivers, C., Rodkey, T., Rojas, A., Rose, M., Rose, R., Ruiz, S.J., Sanders, W., Savary, G., Scherer, S., Scott, G., Shatsman, S., Shen, H., Shetty, J., Shvartsbeyn, A., Sisson, I., Sitter, C.D., Smajs, D., Sneed, A., Sodergren, E., Song, X.-Z., Sorelle, R., Sosa, J., Steinle, M., Strong, R., Sutton, A., Svatek, A., Tabor, P., Taylor, C., Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Usmani, K., Valas, R., Vera, V., Villasana, D., Waldron, L., Walker, B., Wang, J., Wang, Q., Wang, S., Warren, J., Warren, R., Wei, X., White, P., Williams, G., Willson, R., Wleczky, R., Wooden, H., Worley, K., Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V., Yu, F., Zhang, J., Zhou, J., Zhou, X., Zhao, S., Dunn, D., von Niederhausen, A., Weiss, R., Smith, D.R., Holt, R.A., Smith, H.O., Weinstein, G., and Gibbs, R.A.
 Direct Submission
 Unpublished
 2 (bases 1 to 254839)
 Worley, K.C.
 Direct Submission
 Submitted (24-NOV-2001) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
 3 (bases 1 to 254839)
 Rat Genome Sequencing Consortium.
 Direct Submission
 Submitted (13-MAY-2003) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
 On May 13, 2003 this sequence version replaced gi:23265304.
 The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described in the feature table below represents a scaffold in the Atlas assembly (a 'contig-scaffold'). Within each contig-scaffold, individual 'contigs' are ordered and oriented, and separated by sized gaps filled with Ns to the estimated size. The sequence may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.

Center: Baylor College of Medicine
 Center code: BCM
 Web site: http://www.hgsc.bcm.tmc.edu/
 Contact: hgsc-help@bcm.tmc.edu
 ----- Project Information
 Center project name: GJKE
 Center clone name: CH230-98N18
 ----- Summary Statistics
 Assembly program: Atlas 3.0;
 Consensus quality: 241531 bases at least Q40
 Consensus quality: 243998 bases at least Q30
 Consensus quality: 246135 bases at least Q20
 Estimated insert size: 252438; sum-of-contigs estimation
 Quality coverage: 7x in Q20 bases; sum-of-contigs estimation

* NOTE: Estimated insert size may differ from sequence length
 * (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html)
 * NOTE: This sequence may represent more than one clone.
 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 4 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

* 250539: contig of 250539 bp in length
 * 250540 250839: gap of unknown length
 * 250640 251762: contig of 1123 bp in length
 * 251763 251862: gap of unknown length
 * 251863 252935: contig of 1073 bp in length
 * 252936 253035: gap of unknown length
 * 253036 254839: contig of 1804 bp in length.

FEATURES

source

1..254839
 /organism="Rattus norvegicus"
 /mol_type="genomic DNA"
 /db_xref="taxon:10116"
 /clone="CH230-98N18"

misc_feature

1..1230
 /note="wgs end_extension
 clone end:T7"
 3845..5494

misc_feature

/note="wgs end_extension
 clone end:T7"

misc_feature

/note="clone_boundary
 complement(6002..29990)

misc_feature

complement(247726..248554)
 /note="clone_boundary
 clone end:Sp6
 site:EcoRI
 end_sequence:BH311788"

ORIGIN

Alignment Scores:
 Pred. No.: 518 Length: 254839
 Score: 9.00 Matches: 9
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 45.00% Indels: 0
 DB: 2 Gaps: 0

US-10-799-747-116 (1-20) x AC103176 (1-254839)

Qy 5 SerValLeuSerPheLeuLeuTrpThr 13

Db 107271 TCTGTACTTAGTTTCTCTGTGGACT 107245

RESULT 12

AC098074/c

----- Genome Center

LOCUS AC098074 267352 bp DNA linear HTG 10-MAY-2003
 DEFINITION Rattus norvegicus clone CH230-119124, WORKING DRAFT SEQUENCE, 4
 AC098074
 unorderd pieces.
 ACCESSION AC098074
 VERSION GI:30522668
 KEYWORDS HTG; PHASE1; HTGS DRAFT; HTGS_FULLTOP.
 SOURCE Rattus norvegicus (Norway rat)
 ORGANISM Rattus norvegicus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 1 (bases 1 to 267352)
 Muzny, D. Marie., Metzker, M. Lee., Abramson, S., Adams, C., Alder, J.,
 Allen, C., Allen, H., Alsbrooks, S., Amin, A., Anguiano, D.,
 Anyalebechi, V., Aoyagi, A., Ayodeji, M., Baca, E., Baden, H.,
 Baldwin, D., Bandaranaike, D., Barber, M., Barnstead, M., Benahmed, F.,
 Biswal, K., Blair, J., Blankenburg, K., Blyth, P., Brown, M.,
 Bryant, N., Buhay, C., Burch, P., Burrell, K., Calderon, E.,
 Cardenas, V., Carter, K., Cavazos, I., Ceasar, H., Center, A.,
 Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Y., Chen, Z., Chu, J.,
 Cleveland, C., Cockrell, R., Cox, C., Coyle, M., Cree, A., D'Souza, L.,
 Davila, M. L., Davis, C., Davy-Carroll, L., De Anda, C., Dederich, D.,
 Delgado, O., Denson, S., Deramo, C., Ding, Y., Dinh, H., Divya, K.,
 Draper, H., Dugan-Rocha, S., Dunn, A., Durbin, K., Duval, B., Eaves, K.,
 Egan, A., Escoto, M., Eugene, C., Evans, C. A., Falls, T., Fan, G.,
 Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P.,
 Fraser, C. M., Gabisi, A., Ganta, R., Garcia, A., Garner, T., Garza, M.,
 George, G. B., Geer, K., Gill, R., Grady, M., Guerra, W., Guevara, W.,
 Gunaratne, P., Haaland, W., Hamill, C., Hamilton, C., Hamilton, K.,
 Harvey, Y., Havlak, P., Hawes, A., Henderson, N., Hernandez, J.,
 Hernandez, R., Hines, S., Hladun, S. L., Hodgson, A., Hogue, M.,
 Hollins, B., Howells, S., Hulyk, S., Hume, J., Idlebird, D., Jackson, A.,
 Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolivet, A.,
 Karpathy, S., Kelly, S., Kelly, S., Khan, Z., King, L., Kovar, C.,
 Kowis, C., Kraft, C. L., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, J.,
 Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J.,
 Lorensuheva, L., Loulseghe, H., Lozano, R. J., Lu, X., Ma, J.,
 Maheshwari, M., Mahindaratne, M., Mahmoud, M., Malloy, K., Mangum, A.,
 Mangum, B., Mapua, P., Martin, K., Martin, R., Martin, E.,
 Mahoney, S., McLeod, M. P., McNeill, T. Z., Meenen, E.,
 Milosavljevic, A., Miner, G., Minja, E., Montemayor, J., Moore, S.,
 Morgan, M., Morris, K., Morris, S., Munidasa, M., Murphy, M., Nair, L.,
 Nankervis, C., Neal, D., Newton, N., Nguyen, N., Norris, S.,
 Nwakoelam, O., Okunolu, G., Olarnpunsagoon, A., Pal, S., Parks, K.,
 Pasternak, S., Paul, H., Perez, A., Perez, L., Pfankuch, C.,
 Plopper, F., Poindexter, A., Popovic, D., Primus, E., Pu, L.-L.,
 Puazo, M., Quiroz, J., Rachlin, E., Reeves, K., Regier, M. A., Reigh, R.,
 Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, F.,
 Rives, C., Rodkey, T., Rojas, A., Rose, M., Rose, R., Ruiz, S. J.,
 Sanders, W., Savary, G., Scherer, S., Scott, G., Shatsman, S., Shen, H.,
 Shetty, J., Shvartsbeyn, A., Sibson, I., Sitter, C. D., Smajs, D.,
 Sneed, A., Sodergren, E., Song, X.-Z., Sorelle, R., Sosa, J.,
 Steimle, M., Strong, R., Sutton, A., Svatek, A., Tabor, P., Taylor, C.,
 Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Usmani, K.,
 Valas, R., Vera, V., Villasana, D., Waldron, L., Walker, B., Wang, J.,
 Wang, Q., Wang, S., Warren, J., Warren, R., Wei, X., White, F.,
 Williams, G., Willson, R., Wleczyk, R., Wooden, H., Worley, K.,
 Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V.,
 Yu, F., Zhang, J., Zhou, J., Zhou, X., Zhao, S., Dunn, D., von
 Niederhausern, A., Weiss, R., Smith, D. R., Holt, R. A., Smith, H. O.,
 Weinstock, G., and Gibbs, R. A.
 Direct Submission
 Unpublished
 2 (bases 1 to 267352)
 Worley, K. C.
 Direct Submission
 Submitted (23-OCT-2001) Human Genome Sequencing Center, Department
 of Molecular and Human Genetics, Baylor College of Medicine, One
 Baylor Plaza, Houston, TX 77030, USA
 3 (bases 1 to 267352)
 Rat Genome Sequencing Consortium.
 Direct Submission
 Submitted (10-MAY-2003) Human Genome Sequencing Center, Department

of Molecular and Human Genetics, Baylor College of Medicine, One
 Baylor Plaza, Houston, TX 77030, USA
 On May 10, 2003 this sequence version replaced gi:24635703.
 The sequence in this assembly is a combination of BAC based reads
 and whole genome shotgun sequencing reads assembled using Atlas
 (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described
 in the feature table below represents a scaffold in the Atlas
 assembly (a 'contig-scaffold'). Within each contig-scaffold,
 individual sequence contigs are ordered and oriented, and separated
 by sized gaps filled with Ns to the estimated size. The sequence
 may extend beyond the ends of the clone and there may be sequence
 contigs within a contig-scaffold that consist entirely of whole
 genome shotgun sequence reads. Both end sequences and whole genome
 shotgun sequence only contigs will be indicated in the feature
 table.
 ----- Genome Center
 Center: Baylor College of Medicine
 Center code: BCM
 Web site: http://www.hgsc.bcm.tmc.edu/
 Contact: hgsc-help@bcm.tmc.edu
 ----- Project Information
 Center project name: GSGO
 Center clone name: CH230-119124
 ----- Summary Statistics
 Assembly program: Atlas 3.0;
 Consensus quality: 254595 bases at least Q40
 Consensus quality: 257284 bases at least Q30
 Consensus quality: 259179 bases at least Q20
 Estimated insert size: 267168; sum-of-contigs estimation
 Quality coverage: 7x in Q20 bases; sum-of-contigs estimation

 * NOTE: Estimated insert size may differ from sequence length
 (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html)
 * NOTE: This sequence may represent more than one clone.
 * NOTE: This is a 'working draft' sequence. It currently
 consists of 4 contigs. The true order of the pieces
 is not known and their order in this sequence record is
 arbitrary. Gaps between the contigs are represented as
 runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 as soon as it is available and the accession number will
 be preserved.
 * 1 242526: contig of 242526 bp in length
 * 242527 243626: gap of unknown length
 * 242627 264340: contig of 21714 bp in length
 * 264341 264440: gap of unknown length
 * 264441 265515: contig of 1075 bp in length
 * 265516 265615: gap of unknown length
 * 265616 267352: contig of 1737 bp in length.
 FEATURES
 source
 1. 267352
 /organism="Rattus norvegicus"
 /mol_type="genomic DNA"
 /db_xref="taxon:10116"
 /clone="CH230-119124"
 1772. 2630
 /note="clone boundary
 clone end:T7
 site:EcoRI
 end sequence:BH316141"
 18571. 19415
 /note="clone boundary
 clone end:Sp6
 site:EcoRI
 end_sequence:BH316143"
 misc_feature
 misc_feature
 Alignment Scores:
 Pred. No.: 541 Length: 267352
 Score: 9.00 Matches: 9
 Percent Similarity: 100.00% Conservativeness: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 45.00% Indels: 0


```

source
1. .306660
/organism="Rattus norvegicus"
/mol_type="genomic DNA"
/db_xref="taxon:10116"
/clone="CH230-56K13"
1. .1130
/note="wgs end extension"
clone_end:Sp6"
4566. .5998
/note="wgs_end_extension"
clone_end:Sp6"
29817. .30488
/note="clone boundary"
clone_end:Sp6"
site:EcoRI
end_sequence:BH272628"
31694. .33434
/note="wgs_contig"
159947. .161767
/note="wgs_contig"
complement(265403. .266254)
/note="clone boundary"
clone_end:T7
site:EcoRI
end_sequence:BH272626"
272770. .274704
/note="wgs_end_extension"
clone_end:T7"
278461. .280032
/note="wgs_end_extension"
clone_end:T7"

ORIGIN
Alignment Scores:
Pred. No.: 613 Length: 306660
Score: 9.00 Matches: 9
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 45.00% Indels: 0
DB: 2 Gaps: 0

US-10-799-747-116 (1-20) x AC098226 (1-306660)

Qy 6 ValLeuSerPheLeuLeuTrpHrPro 14
|||||
Db 215746 GTGCTTTCCTTCCTGTTATGGACTCCA 215720

RESULT 14
G25673
LOCUS G25673 human STS EST113435, sequence tagged site. 339 bp DNA linear STS 02-JUN-1996
DEFINITION G25673
ACCESSION G25673
VERSION G25673.1 GI:1347905
KEYWORDS STS; STS sequence; primer; sequence tagged site.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
Hudson,T.
1 (bases 1 to 339)
Whitehead Institute/MIT Center for Genome Research; Physically
Mapped STSs
Unpublished (1995)
JOURNAL
COMMENT
Contact: Thomas Hudson
Whitehead Institute/MIT Center for Genome Research
9 Cambridge Center, Cambridge MA 02142 USA
Tel: 617 252 1900
Fax: 617 252 1902
Email: thudson@genome.wi.mit.edu

Primer A: AATTACAGTTTGGTTCTCAGATT
Primer B: CCATGCTAAGTCCAGTGGG

```

```

Primer B: CCATGCTAAGTCCAGTGGG
STS size: 150
PCR Profile:
  Presoak:
    Denaturation:
      Annealing: 56 degrees C
    Polymerization:
      PCR Cycles: 35
    Thermal Cycler:
  Protocol:
    Template: 10 ng
    Primer: each 5 pM
    dNTPs: each 4 nM
    Taq Polymerase: 0.025 units/ul
    Total Vol: 20 ul
  Buffer:
    MgCl2: 1.5 mM
    KCl: 50 mM
    Tris-HCl: 10 mM
    pH: 9.3
  Derived from dbEST (genbank accession T59368).
  Location/Qualifiers
    1..339
      /organism="Homo sapiens"
      /mol_type="genomic DNA"
      /db_xref="taxon:9606"
      /map="527.4 CR from top of Chr2 linkage group"
    1..150
      primer_bind
    1..24
      primer_bind
    complement(131..150)

ORIGIN
Alignment Scores:
Pred. No.: 13.1 Length: 339
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 40.00% Indels: 0
DB: 11 Gaps: 0

US-10-799-747-116 (1-20) x G25673 (1-339)

Qy 3 AlakisserValLeuSerPheLeu 10
|||||
Db 108 GCTCATTCTGTACTTCTCTCTC 131

RESULT 15
G27758
LOCUS G27758 human STS SHGC-33168, sequence tagged site. 356 bp DNA linear STS 29-JUN-1996
DEFINITION G27758
ACCESSION G27758
VERSION G27758.1 GI:1396477
KEYWORDS STS; STS sequence; primer; sequence tagged site.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
Myers,R.M.
1 (bases 1 to 356)
Unpublished (1996)
JOURNAL
COMMENT
Contact: Richard M. Myers
Stanford Human Genome Center (SHGC)
Stanford University School of Medicine
Department of Genetics, M-344, Stanford, CA 94305, USA
Tel: 4157259687
Fax: 4157259689
Email: myers@shgc.stanford.edu

Primer A: AATTACAGTTTGGTTCTCAGATT
Primer B: CCATGCTAAGTCCAGTGGG

```

STS size: 150

PCR Profile:

Initial incubation: 94 degrees C for 90 seconds

Denaturation: 94 degrees C for 15 seconds

Annealing: 62 degrees C for 23 seconds

Polymerization: 72 degrees C for 30 seconds

PCR Cycles: 30

Thermal Cycler: Perkin Elmer 9600

Protocol:

Template: 25 ng

Primer: each 1 uM

dNTPs: each 200 uM

Taq Polymerase: 0.05 units/ul

Total Vol: 10 ul

Buffer:

MgCl2: 2.5 mM

KCl: 50 mM

Tris-HCl: 20 mM

pH: 8.3

Prepared with primer pairs provided by Sandoz, derived from T59368
 -- Washington University/Merck EST sequence.

FEATURES

source

Location/Qualifiers

1..356

/organism="Homo sapiens"

/mol_type="genomic DNA"

/db_xref="taxon:9606"

/map="2"

1..150

STS

primer_bind

primer_bind

complement(131..150)

ORIGIN

Alignment Scores:

Pred. No.:	13.7	Length:	356
Score:	8.00	Matches:	8
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	40.00%	Indels:	0
DB:	11	Gaps:	0

US-10-799-747-116 (1-20) x G27758 (1-356)

QY 3 AlaHisSerValLeuSerPheLeu 10

Db 108 GCTCATTCGTACTTCCTTCCTC 131

Search completed: July 21, 2004, 04:43:02

Job time : 1826 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model
Run on: July 21, 2004, 02:53:21 ; Search time 649 Seconds
(without alignments)
130.915 Million cell updates/sec

Title: US-10-799-747-116
Perfect score: 20
Sequence: 1 MAHSVLSFLWTPVAKSX 20

Scoring table:
Xgapop 60.0 , Xgapext 60.0
Ygapop 60.0 , Ygapext 60.0
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 3373863 seqs, 2124099041 residues

Word size: 1

Total number of hits satisfying chosen parameters: 6742599

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DEV=xlp
-Q=/cgn2_1/USPTO.spool_p/US10799747/runat_19072004_161441_21967/app_query.fasta_1.199
-DB=N_Geneseq_29Jan04 -QMT=fastap -SUFFIX=olig.rng -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=oligo -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THR SCORE=quality -THR MIN=1 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=15
-USER=US10799747 @CGN 1.1 708 @runat_19072004_161441_21967 -NCFU=6 -ICPU=3
-NO.MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=60 -XGAPEXT=60 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=60 -YGAPEXT=60 -DELOP=6 -DELEXT=7

Database :
1: N_Geneseq_29Jan04.*
2: Geneseqn1980s.*
3: Geneseqn1990s.*
4: Geneseqn2000s.*
5: Geneseqn2001as.*
6: Geneseqn2001bs.*
7: Geneseqn2002s.*
8: Geneseqn2003as.*
9: Geneseqn2003bs.*
10: Geneseqn2004s.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	19	95.0	1434	2 AAX37452	Aax37452 Human sec
2	19	95.0	1434	7 ADA39771	Ada39771 Human sec
3	19	95.0	1434	7 ACC50424	Acc50424 Human sec
4	8	40.0	356	6 ABL77790	AbL77790 Human ova
5	8	40.0	1503	7 ADA70841	Ada70841 Rice gene
6	7	35.0	49	7 ACD97100	Acd97100 Human col
7	7	35.0	65	6 ABZ28939	Abz28939 Candida g
8	7	35.0	90	4 AAS23646	Aas23646 Tetracycl

9	7	35.0	109	2 AAH86006	Aah86006 Human sin
10	7	35.0	155	7 ACD97494	Acd97494 Human col
11	7	35.0	263	2 AAT20076	Aat20076 Human gen
12	7	35.0	264	9 ADC93542	Adc93542 E. faeciu
13	7	35.0	293	6 ABN76057	Abn76057 Human ORF
14	7	35.0	294	9 ADC92876	Adc92876 E. faeciu
15	7	35.0	315	2 AAV88416	Aav88416 EST clone
16	7	35.0	352	6 ABQ85808	Abq85808 Arabidops
17	7	35.0	363	5 ABV25515	Abv25515 Human pro
18	7	35.0	377	5 AAF66743	Aaf66743 Novel hum
19	7	35.0	379	8 ACH50509	Ach50509 Human leu
20	7	35.0	394	4 AAK87798	Aak87798 Human dig
21	7	35.0	409	5 ABV38500	Abv38500 Human pro
22	7	35.0	415	3 AAK30357	Aak30357 Human col
23	7	35.0	426	5 ABV08598	Abv08598 Human pro
24	7	35.0	432	6 ABK76313	Abk76313 Bacillus
25	7	35.0	433	7 AEX61964	Aex61964 Arabidops
26	7	35.0	436	4 AAK56191	Aak56191 Human inm
27	7	35.0	440	8 ACH32772	Ach32772 Human end
28	7	35.0	445	8 ACH42384	Ach42384 Human foe
29	7	35.0	455	8 ACH15424	Ach15424 Human adu
30	7	35.0	465	4 AAL11211	Aal11211 Probe #11
31	7	35.0	465	4 AAS2870	Aas2870 Human foe
32	7	35.0	465	4 AAI32473	Aai32473 Probe #11
33	7	35.0	465	4 AAK67520	Aak67520 Human inm
34	7	35.0	465	4 ABA42442	Ab42442 Human bre
35	7	35.0	465	4 ABA22654	Ab22654 Probe #11
36	7	35.0	465	4 AAK26581	Aak26581 Human bon
37	7	35.0	465	4 AAK01123	Aak01123 Human bra
38	7	35.0	465	4 ABS26173	Abs26173 Human liv
39	7	35.0	465	5 AAI01128	Aai01128 Probe #11
40	7	35.0	466	6 ABS01172	Abs01172 Human gen
41	7	35.0	466	6 ABN94789	Abn94789 Gene #128
42	7	35.0	467	8 ACH26710	Ach26710 Human adu
43	7	35.0	468	4 AAK32950	Aak32950 Human bon
44	7	35.0	468	4 AAK07199	Aak07199 Human bra
45	7	35.0	468	4 ABS32679	Abs32679 Human liv

ALIGNMENTS

RESULT 1

AAX37452
ID AAX37452 standard; cDNA; 1434 BP.

XX AAX37452;

DT 06-JUL-1999 (first entry)

XX Human secreted protein cDNA fragment containing gene 2.

XX Human; secreted protein; treatment; prevention; protein therapy; AIDS;
XX gene therapy; diagnosis; cancer; tumor; neurodegenerative disorder;
XX developmental abnormality; fetal deficiency; blood disorder; leukemia;
XX immune system disease; autoimmune disease; hepatic disease; lymphoma;
XX renal disease; inflammation; allergy; Alzheimer's disease; schizophrenia;
XX cognitive disorder; prostate disease; skeletal; cardiac; muscle disorder;
XX pulmonary disorder; transplant rejection; osteoclast; osteoporosis;
XX arthritis; malignancy; digestive; endocrine; infection; ss.

OS Homo sapiens.

XX WO9918208-Al.

PN 15-APR-1999.

XX 01-OCT-1998; 98MO-US020775.

XX 02-OCT-1997; 97US-0060833P.

PR 02-OCT-1997; 97US-0060836P.

PR 02-OCT-1997; 97US-0060837P.

PR 02-OCT-1997; 97US-0060838P.

PR 02-OCT-1997; 97US-0060839P.

PR 02-OCT-1997; 97US-0060843P.
PR 02-OCT-1997; 97US-0060862P.
PR 02-OCT-1997; 97US-0060866P.
PR 02-OCT-1997; 97US-0060874P.
PR 02-OCT-1997; 97US-0060880P.
PR 02-OCT-1997; 97US-0060884P.
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX Duan DR, Florence KA, Rosen CA, Ruben SM, Greene JM, Young P;
PI Ferrie AM, Yu G, Janat F, Ni J, Carter KC, Endress GA, Peng P;
PI Lafleur DW, Shi Y;
XX
XX WPI; 1999-264022/22.
DR P-PSDB; AAY07853.
XX
XX New isolated human genes and the secreted polypeptides they encode.
XX
XX Claim 1a; Page 228; 368pp; English.
XX
XX This invention describes novel isolated human genes and the secreted
XX proteins they encode. The products of the invention are useful for
XX preventing, treating or ameliorating medical conditions, e.g. by protein
XX or gene therapy. Also pathological conditions can be diagnosed by
XX determining the amount of the new polypeptides in a sample or by
XX determining the presence of mutations in the new polynucleotides.
XX Specific uses are described for each of the 101 polynucleotides, based on
XX which tissues they are most highly expressed in, and include developing
XX products for the diagnosis or treatment of cancer, tumours,
XX neurodegenerative disorders, developmental abnormalities and fetal
XX deficiencies, blood disorders, leukemias, diseases of the immune system,
XX autoimmune diseases, hepatic and renal disease, lymphomas, inflammation,
XX allergies, Alzheimer's and cognitive disorders, schizophrenia, prostate
XX disease, skeletal or cardiac muscle disorders, pulmonary disorders,
XX transplant rejection, disorders involving osteoclasts such as
XX osteoporosis, arthritis or malignancies, digestive/endocrine disorders,
XX infections and AIDS. The human secreted proteins of the invention are
XX represented in AAY07852-Y07993 and the encoding nucleic acids are
XX represented in AAX37451-X37552
XX
XX Sequence 1434 BP; 480 A; 204 C; 250 G; 495 T; 0 U; 5 Other;
SQ
Alignment Scores:
Pred. No.: 5,49e-10 Length: 1434
Score: 19.00 Matches: 19
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 95.00% Indels: 0
DB: 2 Gaps: 0
US-10-799-747-116 (1-20) x AAX37452 (1-1434)
QY 1 MetAlaAlaHisSerValLeuSerPheLeuLeuThrProTyrAlaLeuLysSer 19
Db 507 ATGGCAGCCCATTCAGTCTTGAGTTTCTCTCTGGACACCTTATGCTCTGAATCA 563
RESULT 2
ID ADA39771 standard; cDNA; 1434 BP.
XX ADA39771;
AC ADA39771;
XX 20-NOV-2003 (first entry)
DT
XX Human secreted protein encoding cDNA.
DE
XX Human; secreted protein; cancer; hyperproliferative disorder;
XX rheumatoid arthritis; autoimmune disorder; haematopoietic disorder;
KW anaemia; allergic reaction; asthma; cardiovascular disorder;
KW wound healing; cytostatic; immunosuppressive; neutrotropic; neuroprotective;
KW antiviral; antiallergic; hepatotropic; antidiabetic; antiinflammatory;
KW vulnery; cardiant; gene therapy; ss.

OS Homo sapiens.
XX W02002102993-A2.
XX
XX 27-DEC-2002.
XX
XX 19-MAR-2002; 2002WO-US008123.
XX
XX 21-MAR-2001; 2001US-0277340P.
PR 19-JUL-2001; 2001US-0306171P.
PR 13-NOV-2001; 2001US-0331287P.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX Rosen CA, Ruben SM;
PI WPI; 2003-175238/17.
XX
XX New human secreted proteins and nucleic acid molecules, useful for
XX preparing a diagnostic or pharmaceutical composition for diagnosing,
XX preventing or treating cancer or other hyperproliferative disorder,
XX PT asthma, allergies or AIDS.
XX
XX Claim 9; SEQ ID NO 153; 3205pp; English.
XX
XX The invention relates to novel genes ADA39629-ADA40565 and proteins
XX ADA40566-ADA41501 for human secreted proteins, useful for preventing,
XX treating or ameliorating medical conditions e.g. by protein or gene
XX therapy. The polypeptides, nucleic acid molecules, antibodies or their
XX fragments, and agonists or antagonists that bind to the polypeptide are
XX useful for preparing a diagnostic or pharmaceutical composition for
XX diagnosing or treating cancer or other hyperproliferative disorder. The
XX polypeptides and nucleic acid molecules are also useful for detecting,
XX preventing, diagnosing, prognosticating, treating or ameliorating cancer
XX or other hyperproliferative disorders including neoplasms, autoimmune
XX disorders (e.g. diabetes, rheumatoid arthritis, systemic lupus
XX erythematosus, multiple sclerosis, autoimmune thyroiditis or haemolytic
XX anaemia), haematopoietic or haematological disorders (e.g. anaemia,
XX thrombocytopenia), allergic reactions including asthma or eczema,
XX inflammatory disorders (e.g. ischaemia-reperfusion injury, inflammatory
XX bowel disease or Crohn's disease), neurodegenerative disorders (e.g.
XX Alzheimer's disease or Parkinson's disease), cardiovascular disorders
XX (e.g. atherosclerosis, myocarditis), infectious diseases (bacterial,
XX fungal or viral infections including HIV/AIDS), or wound healing and
XX disorders of epithelial cell proliferation. The nucleic acids are also
XX useful for chromosome identification, radiation hybrid mapping or long-
XX range restriction mapping, as molecular weight markers, or as
XX hybridization or diagnostic probes. The polypeptides and antibodies are
XX useful for providing immunological probes for differential identification
XX of the tissues immunohistochemistry assays. Note: The sequence data for
XX this patent did not form part of the printed specification, but was
XX obtained in electronic format directly from WIPO at
XX ftp.wipo.int/pub/published_pct_sequences.
XX
XX Sequence 1434 BP; 480 A; 203 C; 250 G; 496 T; 0 U; 5 Other;
SQ
Alignment Scores:
Pred. No.: 5,49e-10 Length: 1434
Score: 19.00 Matches: 19
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 95.00% Indels: 0
DB: 7 Gaps: 0
US-10-799-747-116 (1-20) x ADA39771 (1-1434)
QY 1 MetAlaAlaHisSerValLeuSerPheLeuLeuThrProTyrAlaLeuLysSer 19
Db 507 ATGGCAGCCCATTCAGTCTTGAGTTTCTCTCTGGACACCTTATGCTCTGAATCA 563
RESULT 3
ACCS0424
ID ACC50424 standard; cDNA; 1434 BP.

XX ACC50424;
XX AC
XX DT 12-JUN-2003 (first entry)
XX DE
XX DE Human secreted protein coding sequence, SEQ ID 91.
XX DE
XX DE Cardiant; antiarrhythmic; antiarteriosclerotic; vasotropic; cytostatic;
KW vulnerary; antiinflammatory; nootropic; neuroprotective;
KW antiparkinsonian; gene therapy; human; cardiovascular disorder; gene; ss.
XX OS
XX OS Homo sapiens.
XX PN WO200295010-A2.
XX PN
XX PD 28-NOV-2002.
XX PF
XX PF 19-MAR-2002; 2002WO-US009785.
XX PR
XX PR 21-MAR-2001; 2001US-0277340P.
XX PR 19-JUL-2001; 2001US-0306171P.
XX PR 13-NOV-2001; 2001US-0331287P.
XX FA (HUMA-) HUMAN GENOME SCI INC.
XX FI
XX FI Rosen CA, Ruben SM;
XX DR WPI; 2003-129429/12.
XX PT Novel human secreted proteins, useful for detecting, preventing,
PT diagnosing, prognosticating, treating and/or ameliorating cardiovascular
PT disorders such as arrhythmia.
XX PS
XX PS Claim 21; SEQ ID NO 91; 1881pp; English.
XX CC The present invention relates to novel human secreted proteins (ABR47633-
CC ABR48145) and their coding sequences (ACC50344-ACC50856). The proteins
CC and their coding sequences are useful for the preparation of a diagnostic
CC or pharmaceutical composition for diagnosing or treating a cardiovascular
CC disorder (e.g., arrhythmia, tachycardia, cardiac arrest, coronary
CC arteriosclerosis and myocardial ischaemia), neural disorders, immune
CC system disorders, muscular disorders, reproductive disorders,
CC gastrointestinal disorders, pulmonary disorders, renal disorders,
CC proliferative disorders and/or cancerous diseases and conditions, for
CC wound healing and epithelial cell proliferation, to treat inflammation or
CC preventing neural damage which occurs in neuronal disorders, for treating or
CC neurodegenerative conditions such as Alzheimer's disease and Parkinson's
CC disease, to enhance bone and periodontal regeneration and aid in tissue
CC transplants or bone grafts, to prevent skin aging or hair loss, to
CC stimulate growth and differentiation of haematopoietic cells and bone
CC marrow cells when used in combination with other cytokines, to maintain
CC organs before transplantation or for supporting cell culture of primary
CC tissues, to increase or decrease differentiation or proliferation of
CC embryonic stem cells, or to modulate mammalian characteristics or
CC metabolism. Note: The sequence data for this patent was published in
CC electronic format and is available from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX SQ Sequence 1434 BP; 480 A; 203 C; 250 G; 496 T; 0 U; 5 Other;

Alignment Scores:
Pred. No.: 5 496-10 Length: 1434
Score: 19.00 Matches: 19
Percent Similarity: 100.00%
Best Local Similarity: 100.00%
Query Match: 95.00%
Conservative: 0
Indels: 0
Gaps: 0
DB: 7

US-10-799-747-116 (1-20) x ACC50424 (1-1434)

QY 1 MetaAlaHisSerValLeuSerPheLeuTyrProTyrAlaLeuLysSer 19
|||||

Db 507 ATGGCAGCCCAATTCAGTCTTGAGTTTCTCTCTGACACCTATGCTCTGAATCA 563
RESULT 4
ABL77790
ID ABL77790 standard; cDNA; 356 BP.
XX AC
XX AC ABL77790;
XX DT 17-MAY-2002 (first entry)
XX DE
XX DE Human ovarian cancer related cDNA clone SEQ ID NO:768.
XX KW Human; ovarian cancer; ovarian tumour; cytostatic; gene; ss.
XX OS Homo sapiens.
XX PN WO200192581-A2.
XX PD 06-DEC-2001.
XX PF 29-MAY-2001; 2001WO-US017756.
XX PR 26-MAY-2000; 2000US-0207484P.
XX FA (CORI-) CORIXA CORP.
XX PI Algate PA, Harlocker SL, Jones R;
XX DR WPI; 2002-122075/16.
XX PT Composition for therapy and diagnosis of ovarian cancer comprising
PT polypeptide of a ovarian tumor polypeptide, polynucleotide encoding
PT polypeptide, antibody specific to polypeptide or T cell expressing
PT polypeptide.
XX PS Claim 1; SEQ ID NO 768; 489pp; English.
XX CC The present invention describes a composition (I) comprising: carriers
CC and immunostimulants; and a polypeptide (II) of a ovarian tumour
CC polypeptide encoded by a polynucleotide (III) having a cDNA sequence (S1)
CC from the 10912 nucleotide sequences as given in ABL77023 to ABL87934,
CC (III) encoding (II) having a sequence (S2), a T cell population of (II),
CC or antigen presenting cells that express (II). (I) has cytostatic
CC activity. An oligonucleotide (IV) that hybridises to (S1) can be used for
CC detecting ovarian cancer in a patient's biological sample preferably
CC serum or ovarian tissue. The method comprises contacting a biological
CC sample from a patient with (IV), detecting the amount of polynucleotide
CC hybridising to (IV) and comparing the amount to a predetermined cutoff
CC value and thereby detecting ovarian cancer in the patient, where the
CC amount of polynucleotide hybridising to (IV) is detected preferably by
CC polymerase chain reaction (PCR). (I) comprising (III) and/or (II) is
CC useful for stimulating and/or expanding T cells specific for an ovarian
CC tumour protein comprising contacting T cells with (III) or (II). (III) is
CC useful in design and preparation of ribozyme molecules for inhibiting
CC expression of the tumour polypeptides and proteins in tumour cells; and
CC to isolate a full length gene from a suitable library e.g., a tumour cDNA
CC library using well known techniques
XX SQ Sequence 356 BP; 57 A; 84 C; 78 G; 136 T; 0 U; 1 Other;

Alignment Scores:
Pred. No.: 19.1 Length: 356
Score: 8.00 Matches: 8
Percent Similarity: 100.00%
Best Local Similarity: 100.00%
Query Match: 40.00%
Conservative: 0
Indels: 0
Gaps: 0
DB: 6

US-10-799-747-116 (1-20) x ABL77790 (1-356)

QY 3 AlaHisSerValLeuSerPheLeu 10
|||||

Db 108 GCTCATTCTGTACTTTCCCTCTC 131

RESULT 5
ADA70841
ID ADA70841 standard; DNA; 1503 BP.
XX
ADA70841;
XX
20-NOV-2003 (first entry)
XX
Rice gene, SEQ ID 4164.
DE
XX Plant; bacterial infection; fungal infection; viral infection; rice;
KW Gene; ds.
XX
XX Oryza sativa.
XX
XX WC2003000898-A1.
XX
XX 03-JAN-2003.
XX
XX 22-JUN-2001; 2001WO-IB001105.
XX
XX 22-JUN-2001; 2001WO-IB001105.
XX
XX (SYGN) SYNGENTA PARTICIPATIONS AG.
XX
XX Chang H, Chen W, Cooper B, Glazebrook J, Goff SA, Hou Y;
PI Katagiri F, Quan S, Tao Y, Whitham S, Xie Z, Zhu T, Zou G;
XX
XX WPI; 2003-175290/17.
XX
XX Identifying at least one gene involved in plant resistance or response to
PT pathogenic infection for conferring resistance or tolerance to a plant to
PT bacterial, fungal or viral infection by determining or detecting plant
PT gene expression.
XX
XX Claim 6; SEQ ID NO 4164; 899pp; English.
XX
XX The present invention relates to a method (M1) for identifying genes
CC involved in plant resistance or response to pathogenic infection. M1
CC comprises identifying a gene whose expression is significantly altered in
CC the incompatible interaction of plant gene expression relative to
CC expression of the gene in an uninfected plant, in a mutant plant that
CC does not express a gene associated with response to pathogenic infection,
CC or in a corresponding incompatible or compatible interaction. (M1) is
CC useful for conferring resistance to resistance or tolerance to a plant to
CC bacterial, fungal or viral infection. The present sequence was used to
XX illustrate the invention.
XX
XX Sequence 1503 BP; 395 A; 329 C; 383 G; 396 T; 0 U; 0 Other;
SQ

Alignment Scores:
Pred. NO.: 72.6 Length: 1503
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 40.00% Indels: 0
DB: 7 Gaps: 0

US-10-799-747-116 (1-20) x ADA70841 (1-1503)
Oy 3 AlatiSerValLeuSerPheLeu 10
Db 266 GCACACAGTGTATCTTTCTT 289

RESULT 6
ACD97100/c
ID ACD97100 standard; cDNA; 49 BP.
XX
XX ACD97100;
XX
XX ACD97100;
XX
XX 23-SEP-2003 (first entry)
DT

Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 35.00% Indels: 0
DB: 7 Gaps: 0

US-10-799-747-116 (1-20) x ACD971100 (1-49)

QY 6 ValLeuSerPheLeuLeuTrp 12
DB 32 GTTTAAGCTTCTCTCTTTTG 12

RESULT 7

ABZ28939
ID ABZ28939 standard; DNA; 65 BP.
XX
AC ABZ28939;
XX
DT 30-JAN-2003 (first entry)
XX
DE Candida gene related tetracycline promoter PCR primer SEQ ID NO 3022.
XX
KW Fungus; yeast; tetracycline; promoter; GRACE strain; biosynthesis;
KW signal transduction; DNA replication; cell division; growth;
KW proliferation; Candida albicans; fungicide; antifungal; PCR; primer; ss.
XX
OS Candida albicans.
XX
FN WO200253728-A2.
XX
PD 11-JUL-2002.
XX
PF 26-DEC-2001; 2001WO-US049486.
XX
PR 29-DEC-2000; 2000US-0259128P.
PR 20-FEB-2001; 2001US-00792024.
PR 22-AUG-2001; 2001US-0314050P.
XX
PA (ELIT-) ELITRA PHARM INC.
XX
PI Roemer T, Jiang B, Boone C, Bussey H, Ohlsen KL;
XX
DR WPI; 2002-566694/60.
XX

Constructing strains for identifying gene products as effective targets for therapeutic intervention, by inactivating in the strain one allele of a gene and placing other allele of the gene under conditional expression.
Claim 76; SEQ ID NO 3022; 167pp + Sequence Listing; English.

The invention relates to constructing (M1) a strain of diploid fungal cells in which both alleles of a gene are modified, comprising modifying one allele by insertion or replacement by a cassette having an expressible selectable marker and modifying other allele by recombination, of a promoter replacement fragment with a heterologous promoter, so that expression of the second allele is regulated by the promoter. (M1) is useful for constructing a strain of diploid fungal cells in which both alleles of a gene are modified. The diploid fungal cells having both alleles modified are useful for identifying a gene that is essential to the survival or growth of a fungus, a gene that contributes to the virulence and/or pathogenicity of a fungus, a gene that contributes to the resistance of a diploid fungus to an antifungal agent, an antifungal agent that inhibits the growth of a diploid fungus and for identifying a therapeutic agent for treatment of a diploid fungus disease. (M1) is useful for identifying a compound which modulates the activity of a gene product, preferably enzymatic activity, carbon compound catabolism, biosynthetic, transporter, transcriptional, translational, signal transduction, DNA replication and cell division activity. The method is useful for identifying a compound having the ability to inhibit growth or proliferation of C. albicans cells and for treating infection by C. albicans. The present sequence is that of a PCR primer used in the method of the invention. Note: The sequence data for this patent is not represented in the printed specification but is based on sequence information supplied to Derwent by the European Patent Office

XX
SQ Sequence 65 BP; 17 A; 12 C; 10 G; 26 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.: 40.5 Length: 65
Score: 7.00 Matches: 7
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 35.00% Indels: 0
DB: 6 Gaps: 0

US-10-799-747-116 (1-20) x ABZ28939 (1-65)

QY 5 SerValLeuSerPheLeuLeu 11
DB 13 TCTGTATTATCATCTTCTTTTG 33

RESULT 8

AAS23646
ID AAS23646 standard; DNA; 90 BP.
XX
AC AAS23646;
XX
DT 04-DEC-2001 (first entry)
XX
DE Tetracycline promoter downstream PCR primer (Tet-Down) #21.
XX
KW Gene identification; essential gene; GRACE; pathogenic fungus;
KW gene replacement and conditional expression; fungal infection;
KW PCR primer; Tet-Down; tetracycline promoter; ss.
XX
OS Escherichia coli.
OS Candida albicans.
XX
PN WO200160975-A2.
XX
PD 23-AUG-2001.
XX
PF 20-FEB-2001; 2001WO-US005551.
XX
PR 18-FEB-2000; 2000US-0183534P.
XX
PA (ELIT-) ELITRA PHARM INC.
XX
PI Roemer T, Jiang B, Boone C, Bussey H;
XX
DR WPI; 2001-489080/53.
XX

Identifying genes essential to fungal metabolisms and identifying potential therapeutic agents that target these genes.

Disclosure; Page 292; 324pp; English.

The present invention relates to novel methods for constructing fungal strains useful for identification and validation of gene products as targets for therapeutic agents, for creating a collection of identified essential genes, and screening assays for the discovery of new drugs. The invention provides the GRACE (gene replacement and conditional expression) method for the construction of mutant organisms referred to as GRACE strains of the organism. The invention can be applied to any organism, particularly a pathogenic fungus e.g. Candida albicans, Aspergillus fumigatus and Cryptococcus neoformans. The methods are useful to identify agents that may be used in the treatment of fungal infections. AAS23626-AAS23686 represent tetracycline promoter downstream PCR primers (Tet-Down) used in the methods of the present invention

SQ Sequence 90 BP; 23 A; 15 C; 17 G; 35 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.: 54.7 Length: 90
Score: 7.00 Matches: 7
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0

Query Match:	35.00%	Indels:	0
DB:	4	Gaps:	0
US-10-799-747-116 (1-20) x AAS23646 (1-90)			
Oy	5 SerValleuSerPheLeuLeu 11		
Db	13 TCTGATTATCATCTTCTATTG 33		
RESULT 9			
AAH86006			
ID	AAH86006 standard; DNA; 109 BP.		
AC			
XX	AAH86006;		
XX			
DT	27-FEB-2002 (first entry)		
XX			
DE	Human single nucleotide polymorphism containing DNA sequence #863.		
XX			
KW	Biallelic marker; polymorphism; human; disease; diagnosis; treatment;		
KW	phenotypic trait; gene therapy; forensic; paternity; mapping; cancer;		
KW	transgenic; single nucleotide polymorphism; SNP; ss.		
XX			
OS	Homo sapiens.		
XX			
EH	Key	Location/Qualifiers	
FT	variation	replace(47,A)	
FT		/*tag= a	
FT		/standard_name= "single nucleotide polymorphism"	
XX			
PN	WO953095-A2.		
XX			
PD	21-OCT-1999.		
XX			
PF	30-MAR-1999; 99WO-US006893.		
XX			
PR	09-APR-1998; 98US-00057871.		
XX			
PA	(WHED) WHITEHEAD INST BIOMEDICAL RES.		
XX			
PI	Lander ES, Wang D, Hudson T;		
XX			
DR	WPI; 1999-620443/53.		
XX			
PT	Polymorphic human genomic sequences and related allele-specific probes		
PT	and primers, useful for genetic analysis, e.g. diagnosis and monitoring		
PT	of disease.		
XX			
PS	Claim 1; Page 123; 330pp; English.		
XX			
CC	This invention describes novel human nucleic acid segments (I) containing		
CC	polymorphic sites. The polynucleotides of (I) are used for, e.g.		
CC	correlating disease polymorphisms (or disease susceptibility) or other		
CC	phenotypic traits (e.g. baldness, obesity, fertility, strength, response		
CC	to drugs etc.); diagnosing and monitoring e.g. cancer, inflammation,		
CC	heart or central nervous system diseases; detecting susceptibility to		
CC	microbial infection; treating or preventing such diseases; forensic		
CC	analysis; gene therapy; paternity testing; mapping genomic loci		
CC	associated with phenotypic traits (and subsequent cloning of the genes		
CC	responsible); and the production of transgenic organisms. Antibodies		
CC	raised against (I) are useful as diagnostic and therapeutic tools and in		
CC	drug screening. AAH85144 - AAH87644 represent the human DNA sequences		
XX	containing biallelic polymorphic sites described in the invention		
XX			
SQ	Sequence 109 BP; 30 A; 28 C; 18 G; 33 T; 0 U; 0 Other;		
Alignment Scores:			
Pred. No.:	65.4	Length:	109
Score:	7.00	Matches:	7
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	35.00%	Indels:	0
DB:	2	Gaps:	0

US-10-799-747-116 (1-20) x AAH86006 (1-109)	
Oy	5 SerValleuSerPheLeuLeu 11
Db	15 AGTGTACTGTCTTTCTACTG 35
RESULT 10	
ACD97494	
ID	ACD97494 standard; cDNA; 155 BP.
XX	
AC	ACD97494;
XX	
DT	23-SEP-2003 (first entry)
XX	
DE	Human colon cancer cell expressed cDNA #5906.
XX	
KW	Open reading frame detection; genome sequencing; colon cancer;
KW	breast cancer; population genome analysis; genetic shift; cancer;
KW	antibiotic resistance; antibiotic non-tolerance; congenital disease;
KW	agriculture; food crop genome; resistance gene; retrovirus;
KW	influenza virus; eukaryotic pathogen detection; trypanosome; Plasmodium;
KW	gene; ss.
XX	
OS	Homo sapiens.
XX	
PN	US2002155438-A1.
XX	
PD	24-OCT-2002.
XX	
PF	27-SEP-1999; 99US-00406117.
XX	
PR	20-NOV-1998; 98US-00196716.
XX	
PA	(SIMP/) SIMPSON A J G.
PA	(NETO/) NETO E D.
PA	(BREN/) BRENTANI R R.
XX	
PI	Simpson AUG, Neto ED, Brentani RR;
XX	
DR	WPI; 2003-182626/18.
XX	
PT	Determining open reading frames of genome of an organism e.g. a human
PT	suffering from cancer involves use of single oligonucleotide primer at
PT	low stringency for preparing single-stranded cDNA from mRNA of
PT	individual.
XX	
PS	Example 9; Page 842; 959pp; English.
XX	
CC	The invention describes a method of determining open reading frames in
CC	the genome of organism, comprising contacting mRNA from cell of organism
CC	with a single oligonucleotide primer (I) at low stringency, preparing
CC	single-stranded cDNA by reverse transcribing mRNA with (I), amplifying
CC	cDNA, sequencing the product, and repeating the contacting, preparing
CC	and amplifying steps with different primers and sequencing resulting
CC	nucleic acids. The method is useful for: determining that a known
CC	nucleotide sequence from a genome of an organism corresponds to a
CC	nucleic acid molecule from a genome of an organism; and for sequencing
CC	all or part of a genome of an organism. mRNA is obtained from mammalian
CC	or human cell which is associated with a pathological condition e.g. a
CC	colon cancer or breast cancer cell. The method is useful for analyses of
CC	populations of subjects and can be used to carry out genetic analyses of
CC	large or small populations. further, it can be used to study living
CC	systems to determine if, e.g. there have been genetic shifts which render
CC	an individual or population more or less likely to be afflicted with
CC	diseases such as cancer, to determine antibiotic resistance or non-
CC	tolerance, and so forth. The method can also be used in the study of
CC	congenital diseases, and the risk of affliction to a foetus, as well as
CC	the study of whether the conditions are likely to be passed to offspring
CC	through ova or sperm. The analyses for pathological conditions can be
CC	carried out in all animals, plants, birds, fish, etc. Using this method,
CC	in the area of agriculture, for example the genomes of food crops can be

CC studied to determine if resistance genes are present, defects in plant
 CC genomes can also be studied in this way. Similarly, the method permits
 CC determination of the pathogens which integrate into the genome, such as
 CC retroviruses and other integrating viruses such as influenza virus, have
 CC undergone shifts or mutations, which may require different approaches to
 CC therapy. This method is also applied to eukaryotic pathogens, such as
 CC trypanosomes, different types of plasmodium, etc. The method essentially
 CC eliminates sequencing of non-coding portions. This sequence represents a
 CC polynucleotide isolated from human colon cancer cell cDNA library

XX SQ Sequence 155 BP; 41 A; 30 C; 33 G; 49 T; 0 U; 2 Other;

Alignment Scores: Pred. No.: 90.6 Length: 155
 Score: 7.00 Matches: 7
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 35.00% Indels: 0
 DB: 7 Gaps: 0

US-10-799-747-116 (1-20) x ACD97494 (1-155)

Qy 5 SerValLeuSerPheLeuLeu 11
 Db 12 AGTGTGCTTCTTCTCTTA 32

RESULT 11

AAAT20076/c
 ID AAAT20076 standard; cDNA to mRNA; 263 BP.

XX AC AAAT20076;

DT 17-JUL-1996 (first entry)

XX Human gene signature HUMGS01218.

XX Gene signature; messenger RNA; mRNA; relative abundance; frequency;
 XX human; cloning; mapping; non-biased library; diagnosis; detection;
 XX cell typing; abnormal cell function; ss.

XX Homo sapiens.

XX WO9514772-A1.

XX 01-JUN-1995.

XX 11-NOV-1994; 94WO-JP001916.

XX 12-NOV-1993; 93JP-00355504.

XX (MATS/) MATSUBARA K.

XX (OKUB/) OKUBO K.

XX Matsubara K, Okubo K;

XX WPI; 1995-206931/27.

XX Single-stranded DNA for identifying gene signatures - isolated from 3'-
 XX directed human cDNA library that reflects relative abundance of corresp.
 XX mRNA in specific human tissues.

XX Claim 1; Page 554; 2245pp; Japanese.

XX A single-stranded DNA (or its complementary strand or the corresp. double
 XX -stranded DNA) which comprises one of the 7837 "GS" sequences given in
 XX AAT19001-T26837 and which is able to hybridise to part of human genomic
 XX DNA, cDNA or mRNA is claimed. The GS (Gene Signature) sequences were
 XX obtained from 3'-directed cDNA libraries prepared from various human
 XX tissues; synthesis of cDNA was initiated from the 3'-end of mRNA by using
 XX poly(T) as the sole primer. Since the 3'- untranslated sequence is unique
 XX to a particular mRNA species, almost all the 3'-oriented cDNAs hybridise
 XX with specific mRNAs. Each library is constructed so as to reflect
 XX accurately the relative abundance of different mRNAs in the particular

CC tissue from which it was derived. The appearance frequency of a given GS
 CC in a cDNA library can be determined (esp. using primers and probes
 CC derived from the GS sequences) as a means of diagnosing abnormal cell
 CC function or for recognising different cell types

XX SQ Sequence 263 BP; 92 A; 34 C; 43 G; 92 T; 0 U; 2 Other;

Alignment Scores: Pred. No.: 148 Length: 263
 Score: 7.00 Matches: 7
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 35.00% Indels: 0
 DB: 2 Gaps: 0

US-10-799-747-116 (1-20) x AAT20076 (1-263)

Qy 5 SerValLeuSerPheLeuLeu 11

Db 82 AGTGTGCTTCTTCTCTC 62

RESULT 12

ADC93542
 ID ADC93542 standard; DNA; 264 BP.

XX AC ADC93542;

XX 01-JAN-2004 (first entry)

XX E. faecium DNA sequence SEQ ID 3169.

XX ds; gene; urinary tract infection; bacteraemia; endocarditis; wound;
 XX abdominal-pelvic infection.

XX Enterococcus faecium.

XX US6583275-B1.

XX 24-JUN-2003.

XX 30-JUN-1998; 98US-00107532.

XX 02-JUL-1997; 97US-0051571P.

XX 14-MAY-1998; 98US-0085598P.

XX (GENO-) GENOME THERAPEUTICS CORP.

XX Doucette-Stamm LA, Bush D;

XX WPI; 2003-799836/75.

XX P-PSDB; ADC97196.

XX New isolated nucleic acid derived from Enterococcus faecium encoding an
 XX Enterococcus faecium polypeptide useful for detection, prevention and
 XX treatment of a pathological condition resulting from a bacterial
 XX infection.

XX Example 1; SEQ ID NO 3169; 243pp; English.

XX The invention relates to an isolated nucleic acid derived from
 XX Enterococcus faecium encoding an Enterococcus faecium polypeptide having
 XX one of 10 fully defined sequences given in the (or comprising 40
 XX sequential nucleotides chosen from any of the nucleic acids, its
 XX complement or sequences hybridising to it). Also included are a
 XX recombinant vector comprising the nucleic acid operably linked to
 XX transcription regulatory element, a cell comprising the vector and a
 XX single-stranded probe comprising the nucleic acid. The nucleic acids are
 XX chosen from 3654 disclosed sequences encoding 3654 disclosed proteins.
 XX The nucleic acids is useful for diagnosing pathological conditions
 XX resulting from E. faecium bacterial infection (e.g. urinary tract
 XX infection, bacteraemia, endocarditis, wounds and abdominal-pelvic
 XX infection) and for screening drugs such as agonists and antagonists. The
 XX nucleic acid is useful for recombinant production of Candida albicans -

CC derived peptides or antisense polypeptides. Pharmaceutical compositions
CC and vaccines containing the nucleic acid are useful for preventing or
CC treating Enterococcus faecium infections. The present sequence represents
CC one of the disclosed E. faecium nucleic acids.

SQ Sequence 264 BP; 67 A; 59 C; 48 G; 88 T; 0 U; 2 Other;

Alignment Scores: 148 Length: 264
Pred. No.: 7.00 Matches: 7
Score: 100.00% Conservative: 0
Percent Similarity: 100.00% Mismatches: 0
Best Local Similarity: 100.00% Indels: 0
Query Match: 35.00% Gaps: 0
DB: 9

US-10-799-747-116 (1-20) x ADC93542 (1-264)

QY 5 SerValLeuSerPheLeuLeu 11
|||||

Db 104 TCGGTTCTGCTTTCTCTTTA 124

RESULT 13

ID ABN76057/c
ID ABN76057 standard; cDNA; 293 BP.

AC ABN76057;

DT 08-JUL-2002 (first entry)

DE Human ORF1004 cDNA, SEQ ID NO:2007.

XX Human; ORF; open reading frame; ORFX; drug screening; diagnosis;
KW disease monitoring; cytokine; cell proliferation; cell differentiation;
KW immune modulation; haematopoiesis regulation; tissue growth;
KW angiogenesis; activin; inhibin; chemotactic; chemokinetic; haemostatic;
KW thrombolytic; tumour inhibition; bodily characteristic; fertility;
KW behaviour; cancer; proliferative disorder; neurological disorder;
KW cardiovascular disease; immune system disorder; organ transplantation;
KW tissue growth disorder; tissue regeneration disorder; diabetes mellitus;
KW hypothyroidism; cholesterol ester storage disease; infection; vulnary;
KW vasotropic; antipsoriatic; antidiabetic; cytoskeletal; neurotropic;
KW neuroprotective; antithrombotic; anticoagulant; thrombolytic;
KW cardiatic; hypertensive; antithyroid; antiinflammatory; immunomodulator;
KW dermatological; analgesic; virucide; antibacterial; fungicide; gene; ss.

XX Homo sapiens.

XX WO200190366-A2.

XX 29-NOV-2001.

XX 24-MAY-2001; 2001WO-US017076.

XX 24-MAY-2000; 2000US-0206690P.

XX (CURA-) CUPAGEN CORP.

XX Leach MD, Shinkets RA;

XX WPI; 2002-106200/14.

XX P-PSDB; ABP32031.

XX Novel human polypeptides and polynucleotides useful for diagnosing,
PT preventing and treating cardiovascular disease, neurodegenerative,
PT hyperproliferative disorders and disorders related to organ
PT transplantation.

XX Claim 1; Page 766; 2508pp; English.

XX Sequences ABP31028-ABP35561 represent 4534 novel human proteins
CC designated ORF (open reading frame) 1-4534, and sequences ABN75054-
CC ABN79587 represent cDNAs encoding them. The invention also encompasses
CC polypeptides at least 80% identical to the ORF1-ORF4534 (collectively

CC referred to as ORFX) proteins, polynucleotides at least 85% identical to
CC the ORFX nucleic acid sequences, vectors and host cells comprising ORFX
CC polynucleotides, the recombinant production of ORFX proteins, antibodies
CC specific for ORFX proteins, methods of detecting ORFX polynucleotides and
CC polypeptides, methods of screening individuals for ORFX expression or
CC activity, and methods of screening individuals for a predisposition to an
CC ORFX-associated disorder. The ORFX proteins of the invention have a wide
CC range of biological activities, such as cytokine, cell proliferation,
CC cell differentiation, immune modulation, haematopoiesis regulation,
CC tissue growth, angiogenesis, activin or inhibin activity, chemotactic/
CC chemokinetic activity, haemostatic activity, thrombolytic activity,
CC receptor/ligand, antiinflammatory activity, tumour inhibition activity,
CC and antiinfective activity, and may also be involved in the determination
CC of bodily characteristics, fertility and behaviour. ORFX proteins,
CC nucleic acids and antibodies may be used in the treatment of cancers,
CC other proliferative disorders such as psoriasis and benign tumours,
CC neurological disorders such as epilepsy and Alzheimer's disease,
CC cardiovascular diseases, immune system disorders, disorders related to
CC organ transplantation, disorders of tissue growth and regeneration,
CC diseases such as diabetes mellitus, hypothyroidism, and cholesterol
CC storage disease, and infectious diseases caused by viral, bacterial,
CC fungal and other pathogens. ORFX nucleic acids may also be used as a
CC source of primers and probes, in the detection of ORFX genomic sequences
CC or transcripts, in the identification and cloning of homologous
CC sequences, in genetic diagnosis, and in forensic biology. The ORFX
CC nucleic acids may additionally be used to produce transgenic animals
CC which may be useful for studying the function and/or activity of ORFX
CC protein, and in drug screening. The ORFX proteins may also be used as
CC immunogens to generate specific antibodies, which are useful in the
CC diagnosis, treatment and monitoring of ORFX-associated diseases

XX Sequence 293 BP; 74 A; 85 C; 66 G; 66 T; 0 U; 2 Other;

Alignment Scores: 163 Length: 293
Pred. No.: 7.00 Matches: 7
Score: 100.00% Conservative: 0
Percent Similarity: 100.00% Mismatches: 0
Best Local Similarity: 100.00% Indels: 0
Query Match: 35.00% Gaps: 0
DB: 6

US-10-799-747-116 (1-20) x ABN76057 (1-293)

QY 5 SerValLeuSerPheLeuLeu 11
|||||

Db 158 AGTGTCTCAGCTTTCTCTT 138

RESULT 14

ADC92876/c

ID ADC92876 standard; DNA; 294 BP.

AC ADC92876;

DT 01-JAN-2004 (first entry)

DE E. faecium DNA sequence SEQ ID 2503.

XX ds; gene; urinary tract infection; bacteraemia; endocarditis; wound;
KW abdominal-pelvic infection.

XX Enterococcus faecium.

XX US6583275-B1.

XX 24-JUN-2003.

XX 30-JUN-1998; 98US-00107532.

XX 02-JUL-1997; 97US-0051571P.

XX 14-MAY-1998; 98US-0085598P.

XX (GENO-) GENOME THERAPEUTICS CORP.

XX


```

PI Doucette-Stamm LA, Bush D;
XX
XX WPI: 2003-799836/75.
DR P-PSDB; ADC96530.
XX
XX New isolated nucleic acid derived from Enterococcus faecium encoding an
PT Enterococcus faecium polypeptide useful for detection, prevention and
PT treatment of a pathological condition resulting from a bacterial
PT infection.
XX
XX Example 1; SEQ ID NO 2503; 243pp; English.
PS
XX The invention relates to an isolated nucleic acid derived from
CC Enterococcus faecium encoding an Enterococcus faecium polypeptide having
CC one of 10 fully defined sequences given in the (or comprising 40
CC sequential nucleotides chosen from any of the nucleic acids, its
CC complement or sequences hybridising to it). Also included are a
CC recombinant vector comprising the nucleic acid operably linked to
CC transcription regulatory element, a cell comprising the vector and a
CC single-stranded probe comprising the nucleic acid. The nucleic acids are
CC chosen from 3654 disclosed sequences encoding 3654 disclosed proteins.
CC The nucleic acids are useful for diagnosing pathological conditions
CC resulting from E. faecium bacterial infection (e.g. urinary tract
CC infection, bacteraemia, endocarditis, wounds and abdominal-pelvic
CC infection) and for screening drugs such as agonists and antagonists. The
CC nucleic acid is useful for recombinant production of Candida albicans -
CC derived peptides or antisense polypeptides. Pharmaceutical compositions
CC and vaccines containing the nucleic acid are useful for preventing or
CC treating Enterococcus faecium infections. The present sequence represents
CC one if the disclosed E. faecium nucleic acids.
XX
XX Sequence 294 BP; 106 A; 57 C; 61 G; 68 T; 0 U; 2 Other;

Alignment Scores:
Pred. No.: 164 Length: 294
Score: 7.00 Matches: 7
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 35.00% Indels: 0
DB: 9 Gaps: 0

US-10-799-747-116 (1-20) x ADC92876 (1-294)

Oy 5 SerValLeuSerPheLeuLeu 11
Db |||||
64 TCGGTTCTGCTTCTTCCTTTTA 44

RESULT 15
AAV88416/c
ID AAV88416 standard; cDNA; 315 BP.
XX
XX AAV88416;
AC
XX
XX 12-FEB-1999 (first entry)
DT
XX
XX EST clone EK581.
DE
XX
XX Expressed sequence tag; secreted protein; haematopoiesis regulator;
KW tissue growth; activin; inhibin; tumour invasion suppressor; EST; human;
KW chemotaxis; chemokinesis; haemostasis; gene therapy; thrombolysis;
KW receptor; ligand; anti-inflammatory; tumour inhibitor; ds.
XX
XX Homo sapiens.
OS
XX
XX WO9845437-A2.
PN
XX
XX 15-OCT-1998.
PD
XX
XX 10-APR-1998; 98WO-US006956.
PF
XX
XX 10-APR-1997; 97US-00837312.
PR
XX
XX (GEMY ) GENETICS INST INC.
PA

```

Jacobs K, McCooy JM, Lavallie ER, Racie LA, Merberg D, Treacy M;
Spaulding V, Agostino MJ;
WPI: 1999-070078/06.

New polynucleotides encoding human secreted proteins - derived from e.g.
human blood, kidney, foetal lung, placenta, testes, brain, ovary,
pituitary, retina and colon cDNA libraries.

Claim 1; Page 389; 64pp; English.

The present sequence represents an expressed sequence tag (EST), and is a
polynucleotide of the invention. The polynucleotides of the invention are
all secreted EST sequences isolated from a variety of human tissue
sources. The EST sequences and proteins encoded by them are predicted to
have useful biological activities which would make them suitable for
treating, preventing or ameliorating medical conditions in humans and
animals, although no supporting data is given. Suggested activities
include nutritional activity, immune stimulating or suppressing activity,
haematopoiesis regulating activity, tissue growth activity, haemostatic
activity/inhibin activity, chemotactic/chemokinetic activity, and thrombolytic activity, receptor/ligand activity, anti-inflammatory
activity, cadherin/tumour invasion suppressor activity, tumour inhibition
activity. The EST sequences are also stated to be useful for gene therapy

Sequence 315 BP; 74 A; 79 C; 65 G; 97 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.: 175 Length: 315
Score: 7.00 Matches: 7
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 35.00% Indels: 0
DB: 2 Gaps: 0

US-10-799-747-116 (1-20) x AAV88416 (1-315)

Oy 2 AlaAlaHisSerValLeuSer 8
|||
Db 215 GCAGCACACAGTGTCCTTTCT 195

Search completed: July 21, 2004, 04:12:23
Job time : 660 secs

This Page Blank (uspto)

GenCore version 5.1.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: July 21, 2004, 03:55:26 ; Search time 72 Seconds
(without alignments)
154.153 Million cell updates/sec

Title: US-10-799-747-116

Perfect score: 20
Sequence: 1 MAHVSFLWTYPALKSX 20

Scoring table:

OLIGO
Xgapop 60.0 , Xgapext 60.0
Ygapop 60.0 , Ygapext 60.0
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 682709 seqs, 277475446 residues

Word size: 1

Total number of hits satisfying chosen parameters: 1359535

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DEV=xlp
-Q/cgn2_1/USPTO_spool_p/US10799747/runat_19072004_161443_22019/app_query.fasta_1.199
-DB=Issued Patents NA -QMT=fastap -SUFFIX=olig.rni -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=oligo -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THR SCORE=quality -THR MIN=1 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10799747@cgn_1_105@runat_19072004_161443_22019 -NCPU=6 -ICPU=3
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=60 -YGAPOP=60 -FPGAPOP=6
-FPGAPEXT=7 -YGAPOP=60 -YGAPEXT=60 -DELOP=6 -DELEXT=7

Database : Issued Patents NA.*

- 1: /cgn2_6/ptodata/2/ina/5A.COMB.seq.*
- 2: /cgn2_6/ptodata/2/ina/5B.COMB.seq.*
- 3: /cgn2_6/ptodata/2/ina/6A.COMB.seq.*
- 4: /cgn2_6/ptodata/2/ina/6B.COMB.seq.*
- 5: /cgn2_6/ptodata/2/ina/PCTUS.COMB.seq.*
- 6: /cgn2_6/ptodata/2/ina/backfiles.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	7	35.0	264	4	US-09-107-532A-3169
2	7	35.0	294	4	US-09-107-532A-2503
3	7	35.0	894	4	US-09-107-532A-1065
4	7	35.0	933	4	US-09-308-003-46
5	7	35.0	1983	1	US-08-221-817-21
6	7	35.0	1983	1	US-08-454-439-21
7	7	35.0	1983	5	PCT-US94-10487-21
8	7	35.0	2056	2	US-08-836-442-1
9	7	35.0	2560	4	US-09-786-256C-29
10	7	35.0	2569	4	US-08-956-171E-74
11	7	35.0	3024	2	US-08-836-943-1
12	7	35.0	3465	4	US-09-134-000C-2717

C 13	7	35.0	3892	3	US-08-569-214-1	Sequence 1, Appli
C 14	7	35.0	3892	3	US-08-937-236-1	Sequence 1, Appli
C 15	7	35.0	4218	4	US-09-081-385-8	Sequence 8, Appli
C 16	7	35.0	34185	4	US-09-545-481-3	Sequence 3, Appli
C 17	7	35.0	99916	4	US-09-816-095-3	Sequence 3, Appli
C 18	7	35.0	536165	4	US-09-214-808-1	Sequence 1, Appli
C 19	7	35.0	786431	4	US-09-751-389-3	Sequence 3, Appli
C 20	7	35.0	1664976	4	US-08-916-421B-1	Sequence 1, Appli
C 21	7	35.0	1664976	4	US-08-916-421B-1	Sequence 1, Appli
C 22	6	30.0	18	4	US-09-308-003-40	Sequence 40, Appli
C 23	6	30.0	18	4	US-09-422-978-9787	Sequence 9787, Ap
C 24	6	30.0	20	1	US-08-271-942A-86	Sequence 86, Appl
C 25	6	30.0	20	3	US-08-779-916A-86	Sequence 86, Appl
C 26	6	30.0	20	5	PCT-US95-08604-86	Sequence 86, Appl
C 27	6	30.0	29	3	US-08-544-381B-152	Sequence 152, App
C 28	6	30.0	32	3	US-08-575-967A-6	Sequence 6, Appli
C 29	6	30.0	35	3	US-09-282-736-16	Sequence 16, Appl
C 30	6	30.0	40	4	US-10-092-246-28	Sequence 28, Appl
C 31	6	30.0	40	4	US-10-150-051-7	Sequence 7, Appli
C 32	6	30.0	40	4	US-10-150-051-8	Sequence 8, Appli
C 33	6	30.0	45	4	US-08-983-564A-21	Sequence 21, Appl
C 34	6	30.0	47	4	US-09-422-978-2157	Sequence 2157, Ap
C 35	6	30.0	65	2	US-03-121-887-5	Sequence 5, Appli
C 36	6	30.0	65	3	US-09-241-353-5	Sequence 5, Appli
C 37	6	30.0	65	3	US-09-245-984-5	Sequence 5, Appli
C 38	6	30.0	65	3	US-09-241-979-5	Sequence 5, Appli
C 39	6	30.0	107	4	US-09-313-294A-1806	Sequence 1806, Ap
C 40	6	30.0	108	1	US-08-441-591-25	Sequence 25, Appl
C 41	6	30.0	108	1	US-08-303-362A-25	Sequence 25, Appl
C 42	6	30.0	108	5	PCT-US95-05600-42	Sequence 42, Appl
C 43	6	30.0	117	3	US-09-284-832-33	Sequence 33, Appl
C 44	6	30.0	120	3	US-09-284-832-32	Sequence 32, Appl
C 45	6	30.0	138	3	US-09-432-020B-1	Sequence 1, Appli

ALIGNMENTS

RESULT 1

US-09-107-532A-3169
; Sequence 3169, Application US/09107532A
; Patent No. 6583275
; GENERAL INFORMATION:

APPLICANT: Lynn A Doucette-Stamm and David Bush
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS
NUMBER OF SEQUENCES: 7310
CORRESPONDENCE ADDRESS:

ADDRESSEE: GENOME THERAPEUTICS CORPORATION
STREET: 100 Beaver Street
CITY: Waltham
STATE: Massachusetts
COUNTRY: USA

ZIP: 02354

COMPUTER READABLE FORM:

MEDIUM TYPE: CD-ROM ISO9660

COMPUTER: PC

OPERATING SYSTEM: <Unknown>

SOFTWARE: ASCII

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/107,532A

FILING DATE: 30-Jun-1998

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 60/085,598

FILING DATE: 14 May 1998

APPLICATION NUMBER: 60/051571

FILING DATE: July 2, 1997

ATTORNEY/AGENT INFORMATION:

NAME: Ariniello, Pamela Deneka

REGISTRATION NUMBER: 40,489

REFERENCE/DOCKET NUMBER: GTC-012

TELECOMMUNICATION INFORMATION:

TELEPHONE: (781)893-5007

TELEFAX: (781)893-8277

INFORMATION FOR SEQ ID NO: 3169:
SEQUENCE CHARACTERISTICS:
LENGTH: 264 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: circular
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Enterococcus faecium
FEATURE:
NAME/KEY: misc feature
LOCATION: (B) LOCATION 1...264
SEQUENCE DESCRIPTION: SEQ ID NO: 3169:
US-09-107-532A-3169
Alignment Scores:
Pred. No.: 15 Length: 264
Score: 7.00 Matches: 7
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 35.00% Indels: 0
DB: 4 Gaps: 0
US-10-799-747-116 (1-20) x US-09-107-532A-3169 (1-264)
QY 5 SerValLeuSerPheLeuLeu 11
DB 104 TCGGTTCTGCTCTTCCTTTA 124
RESULT 2
US-09-107-532A-2503/c
Sequence 2503, Application US/09107532A
Patent No. 6583275
GENERAL INFORMATION:
APPLICANT: Lynn A Doucette-Stamm and David Bush
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS
NUMBER OF SEQUENCES: 7310
CORRESPONDENCE ADDRESS:
ADDRESS: GENOME THERAPEUTICS CORPORATION
STREET: 100 Beaver Street
CITY: Waltham
STATE: Massachusetts
COUNTRY: USA
ZIP: 02354
COMPUTER READABLE FORM:
MEDIUM TYPE: CD/ROM ISO9660
COMPUTER: PC
OPERATING SYSTEM: <Unknown>
SOFTWARE: ASCII
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/107,532A
FILING DATE: 30-Jun-1998
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 60/085,598
FILING DATE: 14 May 1998
APPLICATION NUMBER: 60/051571
FILING DATE: July 2, 1997
ATTORNEY/AGENT INFORMATION:
NAME: Ariniello, Pamela Deneke
REGISTRATION NUMBER: 40,489
REFERENCE/DOCKET NUMBER: GTC-012
TELEPHONE: (781)893-5007
TELEFAX: (781)893-8277
INFORMATION FOR SEQ ID NO: 2503:
SEQUENCE CHARACTERISTICS:
LENGTH: 294 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: circular

MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Enterococcus faecium
FEATURE:
NAME/KEY: misc feature
LOCATION: (B) LOCATION 1...294
SEQUENCE DESCRIPTION: SEQ ID NO: 2503:
US-09-107-532A-2503
Alignment Scores:
Pred. No.: 16.6 Length: 294
Score: 7.00 Matches: 7
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 35.00% Indels: 0
DB: 4 Gaps: 0
US-10-799-747-116 (1-20) x US-09-107-532A-2503 (1-294)
QY 5 SerValLeuSerPheLeuLeu 11
DB 64 TCGGTTCTGCTCTTCCTTTA 44
RESULT 3
US-09-107-532A-1065/c
Sequence 1065, Application US/09107532A
Patent No. 6583275
GENERAL INFORMATION:
APPLICANT: Lynn A Doucette-Stamm and David Bush
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS
NUMBER OF SEQUENCES: 7310
CORRESPONDENCE ADDRESS:
ADDRESS: GENOME THERAPEUTICS CORPORATION
STREET: 100 Beaver Street
CITY: Waltham
STATE: Massachusetts
COUNTRY: USA
ZIP: 02354
COMPUTER READABLE FORM:
MEDIUM TYPE: CD/ROM ISO9660
COMPUTER: PC
OPERATING SYSTEM: <Unknown>
SOFTWARE: ASCII
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/107,532A
FILING DATE: 30-Jun-1998
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 60/085,598
FILING DATE: 14 May 1998
APPLICATION NUMBER: 60/051571
FILING DATE: July 2, 1997
ATTORNEY/AGENT INFORMATION:
NAME: Ariniello, Pamela Deneke
REGISTRATION NUMBER: 40,489
REFERENCE/DOCKET NUMBER: GTC-012
TELEPHONE: (781)893-5007
TELEFAX: (781)893-8277
INFORMATION FOR SEQ ID NO: 1065:
SEQUENCE CHARACTERISTICS:
LENGTH: 894 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: circular
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Enterococcus faecium
FEATURE:

```

; NAME/KEY: misc feature
; LOCATION: (B) LOCATION 1...894
; SEQUENCE DESCRIPTION: SEQ ID NO: 1065:
US-09-107-532A-1065

Alignment Scores:
Pred. No.: 48 Length: 894
Score: 7.00 Matches: 7
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 35.00% Indels: 0
DB: 4 Gaps: 0

US-10-799-747-116 (1-20) x US-09-107-532A-1065 (1-894)

Qy 5 SerValLeuSerPheLeu 11
Db 887 TCTGTTCTTAGCTTTTACTC 867

RESULT 4
US-09-308-003-46/c
; Sequence 46, Application US/09308003
; Patent No. 6326170
; GENERAL INFORMATION:
; APPLICANT: Burnham, Martin K. R.
; APPLICANT: Lonetto, Michael A.
; APPLICANT: Warren, Patrick V.
; TITLE OF INVENTION: NOVEL PROKARYOTIC POLYNUCLEOTIDES,
; TITLE OF INVENTION: POLYPEPTIDES AND THEIR USES
; FILE REFERENCE: GM10093
; CURRENT APPLICATION NUMBER: US/09/308,003
; EARLIER FILING DATE: 1999-05-10
; EARLIER APPLICATION NUMBER: 60/058,710
; EARLIER FILING DATE: 1997-09-12
; NUMBER OF SEQ ID NOS: 52
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 46
; LENGTH: 933
; TYPE: DNA
; ORGANISM: Staphylococcus aureus
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(933)
US-09-308-003-46

Alignment Scores:
Pred. No.: 50 Length: 933
Score: 7.00 Matches: 7
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 35.00% Indels: 0
DB: 4 Gaps: 0

US-10-799-747-116 (1-20) x US-09-308-003-46 (1-933)

Qy 4 HisSerValLeuSerPheLeu 10
Db 891 CACTCTGTTCTCAGCTTCTG 871

RESULT 5
US-08-221-817-21
; Sequence 21, Application US/08221817
; Patent No. 5532151
; GENERAL INFORMATION:
; APPLICANT: Chantry, David
; APPLICANT: Gray, Patrick W.
; APPLICANT: Hoekstra, Merle F.
; TITLE OF INVENTION: A No. 5532151el G Protein-Coupled Receptor
; TITLE OF INVENTION: Kinase GRK6
; NUMBER OF SEQUENCES: 24
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray &
; ADDRESSEE: Borun

```

```

; STREET: 6300 Sears Tower, 233 South Wacker Drive
; CITY: Chicago
; STATE: Illinois
; COUNTRY: USA
; ZIP: 60606
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/221,817
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/123,932
; FILING DATE: 17 SEP 1993
; ATTORNEY/AGENT INFORMATION:
; NAME: No. 5532151and, Greta E.
; REGISTRATION NUMBER: 35,302
; REFERENCE/DOCKET NUMBER: 31981
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (312) 474-6300
; TELEFAX: (312) 474-0448
; TELEX: 25-3856
; INFORMATION FOR SEQ ID NO: 21:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1983 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 13...1740
US-08-221-817-21

Alignment Scores:
Pred. No.: 103 Length: 1983
Score: 7.00 Matches: 7
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 35.00% Indels: 0
DB: 1 Gaps: 0

US-10-799-747-116 (1-20) x US-08-221-817-21 (1-1983)

Qy 3 AlaHisSerValLeuSerPhe 9
Db 1262 GCTCACTCTGTTCTCAGCTTC 1282

RESULT 6
US-08-454-439-21
; Sequence 21, Application US/08454439
; Patent No. 5591618
; GENERAL INFORMATION:
; APPLICANT: Chantry, David
; APPLICANT: Gray, Patrick W.
; APPLICANT: Hoekstra, Merle F.
; TITLE OF INVENTION: A No. 5591618el G Protein-Coupled Receptor
; TITLE OF INVENTION: Kinase GRK6
; NUMBER OF SEQUENCES: 24
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray &
; ADDRESSEE: Borun
; STREET: 6300 Sears Tower, 233 South Wacker Drive
; CITY: Chicago
; STATE: Illinois
; COUNTRY: USA
; ZIP: 60606
; COMPUTER READABLE FORM: disk
; MEDIUM TYPE: Floppy
; COMPUTER: IBM PC compatible

```

```
;
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/454,439
; FILING DATE: 30-MAY-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/221,817
; FILING DATE: 31-MAR-1994
; APPLICATION NUMBER: 08/123,932
; FILING DATE: 17 SEP 1993
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: No. 5591618and, Greta E.
; REGISTRATION NUMBER: 35,302
; REFERENCE/DOCKET NUMBER: 31981
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (312) 474-6300
; TELEFAX: (312) 474-0448
; TELEX: 25-3856
; INFORMATION FOR SEQ ID NO: 21:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1983 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 13...1740
US-08-454-439-21

Alignment Scores:
Pred. No.: 103 Length: 1983
Score: 7.00 Matches: 7
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 35.00% Indels: 0
DB: 1 Gaps: 0

US-10-799-747-116 (1-20) x US-08-454-439-21 (1-1983)
QY 3 AlaHisSerValLeuSerPhe 9
Db 1262 GCTCACTCTGTTCTCAGCTTC 1282

RESULT 7
PCT-US94-10487-21
; Sequence 21, Application PC/TUS9410487
; GENERAL INFORMATION:
; APPLICANT: ICOS Corporation
; TITLE OF INVENTION: A Novel G Protein-Coupled Receptor
; TITLE OF INVENTION: Kinase GRK6
; NUMBER OF SEQUENCES: 24
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray &
; ADDRESSEE: Borun
; STREET: 6300 Sears Tower, 233 South Wacker Drive
; CITY: Chicago
; STATE: Illinois
; COUNTRY: USA
; ZIP: 60606
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US94/10487
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/221,817
```

```
;
; FILING DATE: 31 MAR 1994
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/123,932
; FILING DATE: 17 SEP 1993
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Noland, Greta E.
; REGISTRATION NUMBER: 35,302
; REFERENCE/DOCKET NUMBER: 27866/31981
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (312) 474-6300
; TELEFAX: (312) 474-0448
; TELEX: 25-3856
; INFORMATION FOR SEQ ID NO: 21:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1983 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 13...1740
PCT-US94-10487-21

Alignment Scores:
Pred. No.: 103 Length: 1983
Score: 7.00 Matches: 7
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 35.00% Indels: 0
DB: 5 Gaps: 0

US-10-799-747-116 (1-20) x PCT-US94-10487-21 (1-1983)
QY 3 AlaHisSerValLeuSerPhe 9
Db 1262 GCTCACTCTGTTCTCAGCTTC 1282

RESULT 8
US-08-836-442-1
; Sequence 1, Application US/08836442
; Patent No. 5930293
; GENERAL INFORMATION:
; APPLICANT: DOCHERTY, Andrew, J.P.
; APPLICANT: SLOCOMBE, Patrick, M.
; TITLE OF INVENTION: A HUMAN METALLOPROTEINASE
; TITLE OF INVENTION: VARIANTS THEREOF AND DNA SEQUENCES CODING THEREFOR
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: DIXE, BRONSTEIN, ROBERTS & CUSMAN, LLP
; STREET: 130 Water Street
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/836,442
; FILING DATE: 01-MAY-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/GB96/02181
; FILING DATE: 13-MAR-1997
; APPLICATION NUMBER: GB 9612150.4
; FILING DATE: 11-JUN-1996
; APPLICATION NUMBER: GB 9526229.1
; FILING DATE: 21-DEC-1995
```

```
; APPLICATION NUMBER: GB 9521498.7
; FILING DATE: 20-OCT-1995
; APPLICATION NUMBER: GB 95521495.3
; FILING DATE: 20-OCT-1995
; APPLICATION NUMBER: GB 9518023.8
; FILING DATE: 05-SEP-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Resnick, David S
; REGISTRATION NUMBER: 34,235
; REFERENCE/DOCKET NUMBER: 47425
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617-523-3400
; TELEFAX: 617-523-6440
; TELEX:
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2056 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-08-836-442-1
Alignment Scores:
Pred. No.: 107 Length: 2056
Score: 7.00 Matches: 7
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 35.00% Indels: 0
DB: 2 Gaps: 0
US-10-799-747-116 (1-20) x US-08-836-442-1 (1-2056)
QY 4 HisSerValLeuSerPheLeu 10
Db 1656 CATTCGTCTATCCTTCTTA 1676
RESULT 9
US-09-786-256C-29
; Sequence 29, Application US/09786256C
; Patent No. 6680189
; GENERAL INFORMATION:
; APPLICANT: YOSHIMURA, Koji
; APPLICANT: HIKICHI, Yuichi
; APPLICANT: NISHIMURA, Atsushi
; TITLE OF INVENTION: No. 6680189e1 Protein and DNA Thereof
; FILE REFERENCE: 2544 USOP
; CURRENT APPLICATION NUMBER: US/09/786,256C
; PRIOR FILING DATE: 2001-03-02
; PRIOR APPLICATION NUMBER: PCT/JP99/04766
; PRIOR FILING DATE: 1999-09-02
; PRIOR APPLICATION NUMBER: JP 10-250115
; PRIOR FILING DATE: 1998-09-03
; NUMBER OF SEQ ID NOS: 32
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 29
; LENGTH: 2560
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (55)..(1674)
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)..(2560)
; OTHER INFORMATION: DNA sequence of FIG 1-2 containing SEQ ID NO:3 encoding for prote
; FEATURE:
; NAME/KEY: mat_peptide
; LOCATION: (55)..()
US-09-786-256C-29
Alignment Scores:
Pred. No.: 131 Length: 2560
Score: 7.00 Matches: 2569
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 35.00% Indels: 0
DB: 4 Gaps: 0
US-10-799-747-116 (1-20) x US-08-956-171E-74 (1-2569)
QY 4 HisSerValLeuSerPheLeu 10
Db 1703 CATTCGTCTATCCTTCTTA 1723
RESULT 10
US-08-956-171E-74
; Sequence 74, Application US/08956171E
; Patent No. 6593114
; GENERAL INFORMATION:
; APPLICANT: Charles Kunsch
; Gil H. Choi
; Patrick S. Dillon
; Craig A. Rosen
; Steven C. Barash
; Michael R. Fannon
; TITLE OF INVENTION: Staphylococcus aureus Polynucleotides and Sequences
; NUMBER OF SEQUENCES: 5256
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4mb storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/956,171E
; FILING DATE: 20-Oct-1997
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/009,861
; FILING DATE: January 5, 1996
; APPLICATION NUMBER: 08/781,986
; FILING DATE: January 3, 1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Mark J. Hyman
; REGISTRATION NUMBER: 46,789
; REFERENCE/DOCKET NUMBER: PB248P1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (240) 314-1224
; TELEFAX: (301) 309-8439
; INFORMATION FOR SEQ ID NO: 74:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2569 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; SEQUENCE DESCRIPTION: SEQ ID NO: 74:
US-08-956-171E-74
Alignment Scores:
Pred. No.: 132 Length: 2569
Score: 7.00 Matches: 7
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 35.00% Indels: 0
DB: 4 Gaps: 0
US-10-799-747-116 (1-20) x US-08-956-171E-74 (1-2569)
QY 4 HisSerValLeuSerPheLeu 10
```

```
Db 1609 CATTCTGTTCTCAGCTTCTTG 1629
|||||
RESULT 11
US-08-836-943-1/c
; Sequence 1, Application US/08836943
; Patent No. 5965391
; GENERAL INFORMATION:
; APPLICANT: Reinscheid, Dieter
; APPLICANT: Eikmanns, Bernhard
; APPLICANT: Sahm, Hermann
; TITLE OF INVENTION: DNA WHICH REGULATES GENE EXPRESSION IN
; CORYNEFORM BACTERIA
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: The Firm of Karl F. Ross, PC
; STREET: 5676 Riverdale Ave.
; CITY: Bronx
; STATE: New York
; COUNTRY: USA
; ZIP: 10471
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/836,943
; FILING DATE: 08-MAY-1997
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Myers, Jonathan
; REGISTRATION NUMBER: 26,963
; REFERENCE/DOCKET NUMBER: 20357
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (718) 884-6600
; TELEFAX: 718/601-1099
; TELEX: 620428
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3024 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Corynebacterium glutamicum
; STRAIN: ATCC 13032
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 586..2805
US-08-836-943-1

Alignment Scores:
Pred. No.: 154 Length: 3024
Score: 7.00 Matches: 7
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 35.00% Indels: 0
DB: 2 Gaps: 0

US-10-799-747-116 (1-20) x US-08-836-943-1 (1-3024)
Qy 5 SerValLeuSerPheLeuLeu 11
|||||
Db 2812 AGCGTGCTTAGTTTGTCTT 2792
|||||
RESULT 12
US-09-134-000C-2717/c
; Sequence 2717, Application US/09134000C
; Patent No. 6617156
```

```
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
; FILE REFERENCE: ENTEROCOCCUS FAECALIS FOR DIAGNOSTICS AND THERAPEUTICS
; CURRENT APPLICATION NUMBER: US/09/134,000C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/055,778
; PRIOR FILING DATE: 1997-08-15
; NUMBER OF SEQ ID NOS: 6812
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 2717
; LENGTH: 3465
; TYPE: DNA
; ORGANISM: Enterococcus faecalis
US-09-134-000C-2717

Alignment Scores:
Pred. No.: 176 Length: 3465
Score: 7.00 Matches: 7
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 35.00% Indels: 0
DB: 4 Gaps: 0

US-10-799-747-116 (1-20) x US-09-134-000C-2717 (1-3465)
Qy 5 SerValLeuSerPheLeuLeu 11
|||||
Db 32 TCTGTCTCTCTCTTTT 12
|||||
RESULT 13
US-08-569-214-1/c
; Sequence 1, Application US/08569214
; Patent No. 6165469
; GENERAL INFORMATION:
; APPLICANT: MANN, BARBARA J.
; APPLICANT: PETRI, WILLIAM A.
; TITLE OF INVENTION: RECOMBINANT ENTAMOEBA HISTOLYTICA LECTIN
; TITLE OF INVENTION: SUBUNIT PEPTIDES AND REAGENTS SPECIFIC FOR MEMBERS OF THE
; TITLE OF INVENTION: 170 KD SUBUNIT MULTIGENE FAMILY
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MORRISON & FOERSTER
; STREET: 2000 PENNSYLVANIA AVENUE N.W., STE. 5500
; CITY: WASHINGTON
; STATE: DC
; COUNTRY: USA
; ZIP: 20006-1812
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/569,214
; FILING DATE:
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/US94/06890
; FILING DATE: 17-JUN-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: MURASHIGE, KATE H.
; REGISTRATION NUMBER: 29,959
; REFERENCE/DOCKET NUMBER: 9148-0006.21
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 887-1500
; TELEFAX: (202) 887-0763
; TELEX: 90-4030
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3892 base pairs
; TYPE: nucleic acid
```



```

; STRANDEDNESS: single
; TOPOLOGY: linear
; FEATURE: CDS
; NAME/KEY: join(1..3873, 3877..3882, 3886..3891)
; LOCATION: join(1..3873, 3877..3882, 3886..3891)
US-08-569-214-1
Alignment Scores:
Pred. No.: 196 Length: 3892
Score: 7.00 Matches: 7
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 35.00% Indels: 0
DB: 3 Gaps: 0
US-10-799-747-116 (1-20) x US-08-569-214-1 (1-3892)
QY 4 HisSerValLeuSerPheLeu 10
Db 230 CATTCGTATTATCATTTTAA 210
RESULT 14
US-08-937-236-1/c
; Sequence 1, Application US/08937236
; Patent No. 6187310
; GENERAL INFORMATION:
; APPLICANT: MANN, BARBARA J.
; APPLICANT: PETRI, WILLIAM A.
; APPLICANT: DODSON, JAMES M.
; TITLE OF INVENTION: RECOMBINANT ENTAMORBA HISTOLYTICA LECTIN
; TITLE OF INVENTION: SUBUNIT PEPTIDES AND REAGENTS SPECIFIC FOR MEMBERS OF THE
; TITLE OF INVENTION: 170 KD SUBUNIT MULTIGENE FAMILY
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MORRISON & FOERSTER
; STREET: 2000 PENNSYLVANIA AVENUE N.W., STE. 5500
; CITY: WASHINGTON
; STATE: DC
; COUNTRY: USA
; ZIP: 20006-1812
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/937,236
; FILING DATE:
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/569,214
; FILING DATE: 16 SEPTEMBER 1997
; ATTORNEY/AGENT INFORMATION:
; NAME: LIVNAT, SHMUEL
; REGISTRATION NUMBER: 33,949
; REFERENCE/DOCKET NUMBER: 291482000622
; TELEPHONE: (202) 887-1500
; TELEFAX: (202) 887-0763
; TELEX: 90-4030
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3892 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; FEATURE:
; NAME/KEY: CDS
; LOCATION: join(1..3873, 3877..3882, 3886..3891)
US-08-937-236-1
Alignment Scores:
Pred. No.: 196 Length: 3892

```

```

Score: 7.00 Matches: 7
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 35.00% Indels: 0
DB: 3 Gaps: 0
US-10-799-747-116 (1-20) x US-08-937-236-1 (1-3892)
QY 4 HisSerValLeuSerPheLeu 10
Db 230 CATTCGTATTATCATTTTAA 210
RESULT 15
US-09-081-385-8/c
; Sequence 8, Application US/09081385
; Patent No. 6593456
; GENERAL INFORMATION:
; APPLICANT: Gatanaga, T.
; APPLICANT: Granger, G.A.
; TITLE OF INVENTION: Factors Altering Tumor Necrosis
; TITLE OF INVENTION: Factor Receptor Releasing Enzyme Activity, and Methods
; TITLE OF INVENTION: of Use Thereof
; NUMBER OF SEQUENCES: 154
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MORRISON & FOERSTER
; STREET: 755 PAGE MILL ROAD
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304-1018
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: Windows
; SOFTWARE: FastSeq for Windows Version 2.0b
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/081,385
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/964,747
; FILING DATE: 05-NOV-1997
; APPLICATION NUMBER: 60/030,761
; FILING DATE: 06-NOV-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Wu, Frank
; REGISTRATION NUMBER: 41,396
; REFERENCE/DOCKET NUMBER: 22000-20577.21
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650-813-5600
; TELEFAX: 650-494-0792
; TELEX: 706141
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4218 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: Genomic DNA
US-09-081-385-8
Alignment Scores:
Pred. No.: 212 Length: 4218
Score: 7.00 Matches: 7
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 35.00% Indels: 0
DB: 4 Gaps: 0
US-10-799-747-116 (1-20) x US-09-081-385-8 (1-4218)
QY 5 SerValLeuSerPheLeu 11

```

Db 449 TCAGTGTATCCTTCCTTA 429

Search completed: July 21, 2004, 05:24:13
Job time : 94 secs

GenCore version 5.1.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: July 21, 2004, 04:43:08 ; Search time 3357 Seconds
(without alignments)
29.065 Million cell updates/sec

Title: US-10-799-747-116
Perfect score: 20
Sequence: 1 MAHVSLSFLWTYALKSX 20

Scoring table: OLIGO
Xgapop 60.0 , Xgapext 60.0
Ygapop 60.0 , Ygapext 60.0
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 3191023 seqs, 2439312756 residues

Word size: 1

Total number of hits satisfying chosen parameters: 6371784

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Command line parameters:

-MODE=frame+p2n.model -DEV=xlp
-DB=Published Applications NA -QFMT=fastap -SUFFIX=olig.rnpb -MINMATCH=0.1
-LOOPCL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=oligo
-TRANS=human40.cdi -LIST=45 -DOALIGN=200 -THR SCORE=quality -THR MIN=1
-ALIGN=15 -MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0
-MAXLEN=2000000000 -USER=US10799747@CGN 1 1 723 @runat 19072004 161446 22092
-NCPU=6 -ICPU=3 -NO WMAP -LARGQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100
-LONGLOG -DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=60 -XGAPEXT=60
-FGAPOP=6 -FGAPEXT=7 -YGAPOP=60 -YGAPEXT=60 -DELOP=6 -DELEXT=7

Database :

Published Applications NA:
1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq.*
2: /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq.*
3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq.*
4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq.*
5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq.*
6: /cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq.*
7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq.*
8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq.*
9: /cgn2_6/ptodata/1/pubpna/US09A_PUBCOMB.seq.*
10: /cgn2_6/ptodata/1/pubpna/US09B_PUBCOMB.seq.*
11: /cgn2_6/ptodata/1/pubpna/US09C_PUBCOMB.seq.*
12: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq.*
13: /cgn2_6/ptodata/1/pubpna/US09A_PUBCOMB.seq.*
14: /cgn2_6/ptodata/1/pubpna/US10A_PUBCOMB.seq.*
15: /cgn2_6/ptodata/1/pubpna/US10B_PUBCOMB.seq.*
16: /cgn2_6/ptodata/1/pubpna/US10C_PUBCOMB.seq.*
17: /cgn2_6/ptodata/1/pubpna/US10D_PUBCOMB.seq.*
18: /cgn2_6/ptodata/1/pubpna/US10E_NEW_PUB.seq.*
19: /cgn2_6/ptodata/1/pubpna/US10F_NEW_PUB.seq.*
20: /cgn2_6/ptodata/1/pubpna/US10G_PUBCOMB.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
------------	-------	-------------	--------	-------	-------------

ALIGNMENTS

RESULT 1

US-10-195-730-12
; Sequence 12, Application US/10195730
; Publication No. US2003014492A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et. al
; TITLE OF INVENTION: 101 Human Secreted Proteins
; FILE REFERENCE: P2017P1
; CURRENT APPLICATION NUMBER: US/10/195,730
; CURRENT FILING DATE: 2002-07-16
; PRIOR FILING DATE: 1999-03-31
; PRIOR APPLICATION NUMBER: 60/060,837
; PRIOR FILING DATE: 1997-10-02
; PRIOR APPLICATION NUMBER: 60/060,862
; PRIOR FILING DATE: 1997-10-02
; NUMBER OF SEQ ID NOS: 390
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 12
; LENGTH: 1434
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-195-730-12

1	19	95.0	1434	15	US-10-195-730-12	Sequence 12, Appl
2	8	40.0	356	9	US-09-867-701-768	Sequence 768, App
C 3	8	40.0	670	13	US-10-027-632-187894	Sequence 187894, A
C 4	8	40.0	670	16	US-10-027-632-187894	Sequence 187894, A
C 5	8	40.0	740	13	US-10-027-632-11891	Sequence 11891, A
C 6	8	40.0	740	16	US-10-027-632-11891	Sequence 11891, A
C 7	8	40.0	1627	17	US-10-437-963-82948	Sequence 82948, A
C 8	8	40.0	2037	17	US-10-437-963-82948	Sequence 82948, A
C 9	8	40.0	6681	17	US-10-322-281-411	Sequence 411, App
C 10	8	40.0	133787	17	US-10-322-281-858	Sequence 858, App
C 11	8	40.0	196063	17	US-10-322-281-612	Sequence 612, App
C 12	8	40.0	1601042	16	US-10-027-632-59064	Sequence 59064, A
C 13	8	40.0	1601042	16	US-10-027-632-59064	Sequence 59064, A
C 14	7	35.0	65	15	US-10-032-585-3022	Sequence 3022, App
C 15	7	35.0	163	13	US-10-424-599-127928	Sequence 127928, A
C 16	7	35.0	191	13	US-10-424-599-98581	Sequence 98581, A
C 17	7	35.0	221	13	US-10-085-783A-36161	Sequence 36161, A
C 18	7	35.0	221	16	US-10-242-535A-36161	Sequence 36161, A
C 19	7	35.0	240	13	US-10-424-599-132181	Sequence 132181, A
C 20	7	35.0	279	13	US-10-424-599-22323	Sequence 22323, A
C 21	7	35.0	292	13	US-10-424-599-133734	Sequence 133734, A
C 22	7	35.0	293	11	US-09-864-408A-2007	Sequence 2007, App
C 23	7	35.0	315	14	US-10-040-739-894	Sequence 894, App
C 24	7	35.0	324	13	US-10-027-632-143650	Sequence 143650, A
C 25	7	35.0	324	16	US-10-027-632-143650	Sequence 143650, A
C 26	7	35.0	348	17	US-10-437-963-77562	Sequence 77562, A
C 27	7	35.0	352	9	US-09-770-791-678	Sequence 678, App
C 28	7	35.0	379	10	US-09-918-995-37721	Sequence 37721, A
C 29	7	35.0	415	17	US-10-437-963-38874	Sequence 38874, A
C 30	7	35.0	423	13	US-10-027-632-104598	Sequence 104598, A
C 31	7	35.0	423	16	US-10-027-632-298016	Sequence 298016, A
C 32	7	35.0	423	13	US-10-027-632-298016	Sequence 298016, A
C 33	7	35.0	423	16	US-10-027-632-298016	Sequence 298016, A
C 34	7	35.0	431	17	US-10-437-963-44015	Sequence 44015, A
C 35	7	35.0	432	9	US-09-974-300-3604	Sequence 3604, App
C 36	7	35.0	433	13	US-09-770-423-79	Sequence 79, Appl
C 37	7	35.0	435	13	US-10-424-599-41140	Sequence 41140, A
C 38	7	35.0	440	10	US-09-918-995-19984	Sequence 19984, A
C 39	7	35.0	442	13	US-10-424-599-59524	Sequence 59524, A
C 40	7	35.0	445	10	US-09-918-995-29596	Sequence 29596, A
C 41	7	35.0	455	13	US-10-027-632-179616	Sequence 179616, A
C 42	7	35.0	455	16	US-10-027-632-179616	Sequence 179616, A
C 43	7	35.0	455	9	US-09-864-761-1120	Sequence 1120, App
C 44	7	35.0	465	9	US-09-796-692-7472	Sequence 7472, App
C 45	7	35.0	465	9	US-09-796-692-7472	Sequence 7472, App

```
Alignment Scores:
Pred. No.: 1.22e-10 Length: 1434
Score: 19.00 Matches: 19
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 95.00% Indels: 0
DB: 15 Gaps: 0

US-10-799-747-116 (1-20) x US-10-195-730-12 (1-1434)

Qy 1 MetAlaHisSerValLeuSerPheLeuLeuTrpThrProTyAlaLeuLysSer 19
Db 507 ATGCAGCCCATTCAGTCTTGAGTTTCTTCTCTGGACACCTTATGCTCTGAATCA 563

RESULT 2
US-09-867-701-768
; Sequence 768, Application US/09867701
; Patent No. US20020132237A1
; GENERAL INFORMATION:
; APPLICANT: Aglate, Paul A.
; APPLICANT: Jones, Robert
; APPLICANT: Harlocker, Susan L.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; TITLE OF INVENTION: AND DIAGNOSIS OF OVARIAN CANCER
; FILE REFERENCE: 210121.497
; CURRENT APPLICATION NUMBER: US/09/867,701
; CURRENT FILING DATE: 2001-05-29
; NUMBER OF SEQ ID NOS: 10912
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 768
; LENGTH: 356
; TYPE: DNA
; ORGANISM: Homo sapien
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(356)
; OTHER INFORMATION: n = A,T,C or G
US-09-867-701-768

Alignment Scores:
Pred. No.: 11.2 Length: 356
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 40.00% Indels: 0
DB: 9 Gaps: 0

US-10-799-747-116 (1-20) x US-09-867-701-768 (1-356)

Qy 3 AlaHisSerValLeuSerPheLeu 10
Db 108 GTCATTCGTACTTCTCTCTC 131

RESULT 3
US-10-027-632-187894/c
; Sequence 187894, Application US/10027632
; Publication No. US20020198371A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; TITLE OF INVENTION: Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR FILING DATE: 2000-04-20
; PRIOR FILING DATE: 2000-03-29
; PRIOR FILING DATE: 2000-02-24
; PRIOR FILING DATE: 1999-11-23
; PRIOR FILING DATE: 1999-09-28
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 187894
; LENGTH: 670
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(670)
; OTHER INFORMATION: n = A,T,C or G
US-10-027-632-187894

Alignment Scores:
Pred. No.: 20 Length: 670
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
```

```
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 187894
; LENGTH: 670
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(670)
; OTHER INFORMATION: n = A,T,C or G
US-10-027-632-187894

Alignment Scores:
Pred. No.: 20 Length: 670
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 40.00% Indels: 0
DB: 13 Gaps: 0

US-10-799-747-116 (1-20) x US-10-027-632-187894 (1-670)

Qy 1 MetAlaHisSerValLeuSer 8
Db 36 ATGCAGCCCATAGTGTGCTCTCC 13

RESULT 4
US-10-027-632-187894/c
; Sequence 187894, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; TITLE OF INVENTION: Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR FILING DATE: 2000-03-29
; PRIOR FILING DATE: 2000-02-24
; PRIOR FILING DATE: 1999-11-23
; PRIOR FILING DATE: 1999-09-28
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 187894
; LENGTH: 670
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(670)
; OTHER INFORMATION: n = A,T,C or G
US-10-027-632-187894

Alignment Scores:
Pred. No.: 20 Length: 670
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
```

Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 40.00% Indels: 0
 DB: 16 Gaps: 0

US-10-799-747-116 (1-20) x US-10-027-632-187894 (1-670)

Qy 1 MetAlaAlaHisSerValLeuSer 8
 Db 36 ATGGCAGCCCATAGTGTCTCTCC 13

RESULT 5

US-10-027-632-11891/c
 ; Sequence 11891, Application US/10027632
 ; Publication No. US20020198371A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Wang, David G.
 ; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
 ; FILE REFERENCE: 108827.129
 ; CURRENT APPLICATION NUMBER: US/10/027,632
 ; CURRENT FILING DATE: 2002-04-30
 ; PRIOR APPLICATION NUMBER: US 60/218,006
 ; PRIOR FILING DATE: 2000-07-12
 ; PRIOR APPLICATION NUMBER: US 60/198,676
 ; PRIOR FILING DATE: 2000-04-20
 ; PRIOR APPLICATION NUMBER: US 60/193,483
 ; PRIOR FILING DATE: 2000-03-29
 ; PRIOR APPLICATION NUMBER: US 60/185,218
 ; PRIOR FILING DATE: 2000-02-24
 ; PRIOR APPLICATION NUMBER: US 60/167,363
 ; PRIOR FILING DATE: 1999-11-23
 ; PRIOR APPLICATION NUMBER: US 60/156,358
 ; PRIOR FILING DATE: 1999-09-28
 ; PRIOR APPLICATION NUMBER: US 60/146,002
 ; PRIOR FILING DATE: 1999-08-09
 ; NUMBER OF SEQ ID NOS: 325720
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 11891
 ; LENGTH: 740
 ; TYPE: DNA
 ; ORGANISM: Human
 ; US-10-027-632-11891

Alignment Scores:

Pred. No.: 21.9 Length: 740
 Score: 8.00 Matches: 8
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 40.00% Indels: 0
 DB: 13 Gaps: 0

US-10-799-747-116 (1-20) x US-10-027-632-11891 (1-740)

Qy 5 SerValLeuSerPheLeuLeuTrp 12
 Db 266 AGTGTCTTAAGCTTCCTCTCTGG 243

RESULT 6

US-10-027-632-11891/c
 ; Sequence 11891, Application US/10027632
 ; Publication No. US20030204075A9
 ; GENERAL INFORMATION:
 ; APPLICANT: Wang, David G.
 ; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
 ; FILE REFERENCE: 108827.129
 ; CURRENT APPLICATION NUMBER: US/10/027,632
 ; CURRENT FILING DATE: 2002-04-30
 ; PRIOR APPLICATION NUMBER: US 60/218,006
 ; PRIOR FILING DATE: 2000-07-12
 ; PRIOR APPLICATION NUMBER: US 60/198,676
 ; PRIOR FILING DATE: 2000-04-20
 ; PRIOR APPLICATION NUMBER: US 60/193,483

; PRIOR FILING DATE: 2000-03-29
 ; PRIOR APPLICATION NUMBER: US 60/185,218
 ; PRIOR FILING DATE: 2000-02-24
 ; PRIOR APPLICATION NUMBER: US 60/167,363
 ; PRIOR FILING DATE: 1999-11-23
 ; PRIOR APPLICATION NUMBER: US 60/156,358
 ; PRIOR FILING DATE: 1999-09-28
 ; PRIOR APPLICATION NUMBER: US 60/146,002
 ; PRIOR FILING DATE: 1999-08-09
 ; NUMBER OF SEQ ID NOS: 325720
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 11891
 ; LENGTH: 740
 ; TYPE: DNA
 ; ORGANISM: Human
 ; US-10-027-632-11891

Alignment Scores:

Pred. No.: 21.9 Length: 740
 Score: 8.00 Matches: 8
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 40.00% Indels: 0
 DB: 16 Gaps: 0

US-10-799-747-116 (1-20) x US-10-027-632-11891 (1-740)

Qy 5 SerValLeuSerPheLeuLeuTrp 12
 Db 266 AGTGTCTTAAGCTTCCTCTCTGG 243

RESULT 7

US-10-437-963-82948
 ; Sequence 82948, Application US/10437963
 ; Publication No. US20040123343A1
 ; GENERAL INFORMATION:
 ; APPLICANT: La Rosa, Thomas J.
 ; APPLICANT: Kovalic, David K.
 ; APPLICANT: Zhou, Yihua
 ; APPLICANT: Cao, Yongwei
 ; APPLICANT: Wu, Wei
 ; APPLICANT: Boukharov, Andrey A.
 ; APPLICANT: Barbazuk, Brad
 ; APPLICANT: Li, Ping
 ; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
 ; FILE REFERENCE: 38-21(53221)B
 ; CURRENT APPLICATION NUMBER: US/10/437,963
 ; CURRENT FILING DATE: 2003-05-14
 ; NUMBER OF SEQ ID NOS: 204966
 ; SEQ ID NO 82948
 ; LENGTH: 1627
 ; TYPE: DNA
 ; ORGANISM: Oryza sativa
 ; FEATURE:
 ; OTHER INFORMATION: Clone ID: PAT_MRT4530_82327C.1
 ; US-10-437-963-82948

Alignment Scores:

Pred. No.: 45.2 Length: 1627
 Score: 8.00 Matches: 8
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 40.00% Indels: 0
 DB: 17 Gaps: 0

US-10-799-747-116 (1-20) x US-10-437-963-82948 (1-1627)

Qy 3 AlaHisSerValLeuSerPheLeu 10
 Db 266 GCACACAGTGTTCCTCTCTCT 289

RESULT 8

US-10-437-963-82916
; Sequence 82916, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 82916
; LENGTH: 2037
; TYPE: DNA
; ORGANISM: Oryza sativa
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MKT4530_82299C.1
US-10-437-963-82916
Alignment Scores:
Pred. No.: 55.6 Length: 2037
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 40.00% Indels: 0
DB: 17 Gaps: 0
US-10-799-747-116 (1-20) x US-10-437-963-82916 (1-2037)
QY 6 ValLeuSerPheLeuThr 13
DB 69 GTGCTATCCTTCTCTCTGGACA 92
RESULT 9
US-10-322-281-411
; Sequence 411, Application US/10322281
; Publication No. US20040126762A1
; GENERAL INFORMATION:
; APPLICANT: David W. Morris
; APPLICANT: Marc S. Malandro
; TITLE OF INVENTION: Novel Compositions and Methods in Cancer
; FILE REFERENCE: 529452001000
; CURRENT APPLICATION NUMBER: US/10/322,281
; CURRENT FILING DATE: 2002-12-17
; NUMBER OF SEQ ID NOS: 866
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 411
; LENGTH: 66681
; TYPE: DNA
; ORGANISM: Mus musculus
; FEATURE:
; NAME/KEY: misc.feature
; LOCATION: (1)...(66681)
; OTHER INFORMATION: n = A,T,C or G
US-10-322-281-411
Alignment Scores:
Pred. No.: 1.37e+03 Length: 66681
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 40.00% Indels: 0
DB: 17 Gaps: 0
US-10-799-747-116 (1-20) x US-10-322-281-411 (1-66681)

QY 4 HisSerValLeuSerPheLeu 11
DB 38196 CACTCTGTCTTCATTCCTTTTG 38219
RESULT 10
US-10-322-281-858/c
; Sequence 858, Application US/10322281
; Publication No. US20040126762A1
; GENERAL INFORMATION:
; APPLICANT: David W. Morris
; APPLICANT: Marc S. Malandro
; TITLE OF INVENTION: Novel Compositions and Methods in Cancer
; FILE REFERENCE: 529452001000
; CURRENT APPLICATION NUMBER: US/10/322,281
; CURRENT FILING DATE: 2002-12-17
; NUMBER OF SEQ ID NOS: 866
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 858
; LENGTH: 133787
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc.feature
; LOCATION: (1)...(133787)
; OTHER INFORMATION: n = A,T,C or G
US-10-322-281-858
Alignment Scores:
Pred. No.: 2.61e+03 Length: 133787
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 40.00% Indels: 0
DB: 17 Gaps: 0
US-10-799-747-116 (1-20) x US-10-322-281-858 (1-133787)
QY 4 HisSerValLeuSerPheLeu 11
DB 35487 CACTCAGTTCATCATTCCTTCTT 35464
RESULT 11
US-10-322-281-612
; Sequence 612, Application US/10322281
; Publication No. US20040126762A1
; GENERAL INFORMATION:
; APPLICANT: David W. Morris
; APPLICANT: Marc S. Malandro
; TITLE OF INVENTION: Novel Compositions and Methods in Cancer
; FILE REFERENCE: 529452001000
; CURRENT APPLICATION NUMBER: US/10/322,281
; CURRENT FILING DATE: 2002-12-17
; NUMBER OF SEQ ID NOS: 866
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 612
; LENGTH: 196063
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc.feature
; LOCATION: (1)...(196063)
; OTHER INFORMATION: n = A,T,C or G
US-10-322-281-612
Alignment Scores:
Pred. No.: 3.7e+03 Length: 196063
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 40.00% Indels: 0
DB: 17 Gaps: 0
US-10-799-747-116 (1-20) x US-10-322-281-612 (1-196063)

```

QY      3 AlahisServalLeuSerPheLeu 10
Db      50669 GCTCATCTGTACTTTCCTTCCTC 50692

RESULT 12
US-10-027-632-59064/c
; Sequence 59064, Application US/10027632
; Publication No. US20020198371A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; PRIOR FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 59064
; LENGTH: 1601042
; TYPE: DNA
; ORGANISM: Human
; NAME/KEY: misc feature
; LOCATION: (1)...(1601042)
; OTHER INFORMATION: n = A,T,C or G
US-10-027-632-59064

Alignment Scores:
Pred. No.:      2.55e+04      Length:      1601042
Score:          8.00         Matches:      8
Percent Similarity: 100.00%   Conservative: 0
Best Local Similarity: 100.00% Mismatches:      0
Query Match:    40.00%       Indels:        0
DB:             13          Gaps:          0

US-10-799-747-116 (1-20) x US-10-027-632-59064 (1-1601042)

QY      3 AlahisServalLeuSerPheLeu 10
Db      638163 GCCCACTCGGTCTTAGCTTCTG 638140

RESULT 13
US-10-027-632-59064/c
; Sequence 59064, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; PRIOR FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29

```

```

; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 59064
; LENGTH: 1601042
; TYPE: DNA
; ORGANISM: Human
; NAME/KEY: misc feature
; LOCATION: (1)...(1601042)
; OTHER INFORMATION: n = A,T,C or G
US-10-027-632-59064

Alignment Scores:
Pred. No.:      2.55e+04      Length:      1601042
Score:          8.00         Matches:      8
Percent Similarity: 100.00%   Conservative: 0
Best Local Similarity: 100.00% Mismatches:      0
Query Match:    40.00%       Indels:        0
DB:             16          Gaps:          0

US-10-799-747-116 (1-20) x US-10-027-632-59064 (1-1601042)

QY      3 AlahisServalLeuSerPheLeu 10
Db      638163 GCCCACTCGGTCTTAGCTTCTG 638140

RESULT 14
US-10-032-585-3022
; Sequence 3022, Application US/10032585
; Publication No. US20030180953A1
; GENERAL INFORMATION:
; APPLICANT: Terry, Roemer D.
; APPLICANT: Bo, Jiang
; APPLICANT: Charles, Boone
; APPLICANT: Howard, Bussey
; TITLE OF INVENTION: Gene Disruption Methodologies for Drug Target Discovery
; FILE REFERENCE: 10182-005-999
; CURRENT APPLICATION NUMBER: US/10/032,585
; CURRENT FILING DATE: 2001-12-20
; NUMBER OF SEQ ID NOS: 8000
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 3022
; LENGTH: 65
; TYPE: DNA
; ORGANISM: Candida albicans
US-10-032-585-3022

Alignment Scores:
Pred. No.:      26.1         Length:      65
Score:          7.00         Matches:      7
Percent Similarity: 100.00%   Conservative: 0
Best Local Similarity: 100.00% Mismatches:      0
Query Match:    35.00%       Indels:        0
DB:             15          Gaps:          0

US-10-799-747-116 (1-20) x US-10-032-585-3022 (1-65)

QY      5 ServalLeuSerPheLeu 11
Db      13 TCTGTATTATCATCTTATTG 33

RESULT 15
US-10-424-599-127928
; Sequence 127928, Application US/10424599
; Publication No. US20040031072A1

```

```
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 127928
; LENGTH: 163
; TYPE: DNA
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_86523C.1
US-10-424-599-127928

Alignment Scores:
Pred. No.:      60.8      Length:      163
Score:          7.00      Matches:      7
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match:      35.00% Indels:      0
DB:               13      Gaps:      0

US-10-799-747-116 (1-20) x US-10-424-599-127928 (1-163)

Qy      4 HisSerValLeuSerPheLeu 10
        |||||
Db      40 CACTCTGTGTGTCCTTTCTG 60
```

Search completed: July 21, 2004, 07:54:36
Job time : 3580 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: July 21, 2004, 03:49:02 ; Search time 2467 Seconds
(without alignments)
242.093 Million cell updates/sec

Title: US-10-799-747-116
Perfect score: 20
Sequence: 1 MAHSVLSFLWTFYALKSX 20

Scoring table: OLIGO
Xgapop 60.0 , Xgapext 60.0
Ygapop 60.0 , Ygapext 60.0
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 27513289 seqs, 14931090276 residues

Word size: 1

Total number of hits satisfying chosen parameters: 55023894

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Command line parameters:

-MODEL=frame+p2n.model -DEV=xlp
-Q=/cgm2_1/USPTO.spool_p/US10799747/runat_19072004.161443_21994/app_query.fasta_1.199
-DB=EST -QFMT=fastcap -SUFFIX=olig.rst -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=oligo -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR SCORE=quality -THR_MIN=1 -ALIGN=15 -MODE=LOCAL -OUTFMT=pto
-NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10799747@cgm2_1_5180 @runat_19072004.161443_21994 -NCPU=6 -ICPU=3
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=60 -XGAPEXT=60 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=60 -YGAPEXT=60 -DELOP=6 -DELEXT=7

Database :

EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_est1:*
10: gb_est2:*
11: gb_htc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: em_gss_hum:*
18: em_gss_inv:*
19: em_gss_pln:*
20: em_gss_vrt:*
21: em_gss_fun:*
22: em_gss_man:*
23: em_gss_mus:*
24: em_gss_pro:*
25: em_gss_rod:*
26: em_gss_phg:*
27: em_gss_vrl:*
28: gb_gss1:*

29: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
C 1	19	95.0	240	10	BF910533	BF910533 CM4-UT004
2	19	95.0	575	12	BM722991	BM722991 UI-E-E01
3	19	95.0	765	12	BI914473	BI914473 603182264
4	19	95.0	785	12	BG484396	BG484396 602505037
5	19	95.0	1201	13	BX441923	BX441923 BX441923
6	10	50.0	621	9	AL035941	AL035941 DKF2p564B
7	9	45.0	496	28	AQ293004	AQ293004 HS 2225 A
8	8	40.0	239	9	AV851019	AV851019 AV851019
9	8	40.0	272	12	BG371993	BG371993 UI-R-CV0
10	8	40.0	290	29	CE488576	CE488576 tigr-ges-
11	8	40.0	323	28	AQ350287	AQ350287 RPC111-13
12	8	40.0	328	29	CC808823	CC808823 ZMMB6C047
13	8	40.0	336	14	CD202965	CD202965 MS1-0139P
14	8	40.0	356	14	T59368	T59368 yb57h01.s1
15	8	40.0	385	29	CE286752	CE286752 tigr-ges-
16	8	40.0	396	28	BH363737	BH363737 CH230-127
17	8	40.0	400	28	AQ475321	AQ475321 CITBI-E1-
18	8	40.0	419	13	BW204790	BW204790 BW204790
19	8	40.0	430	14	CD201054	CD201054 MS1-0124U
20	8	40.0	430	14	CD201131	CD201131 MS1-0124U
21	8	40.0	441	9	AV888988	AV888988 AV888988
22	8	40.0	444	10	BF412147	BF412147 UI-R-BF1-
23	8	40.0	448	14	CD113088	CD113088 ME1-0028T
24	8	40.0	448	14	CD201074	CD201074 MS1-0124U
25	8	40.0	448	14	CD201080	CD201080 MS1-0124U
26	8	40.0	448	14	CD201099	CD201099 MS1-0124U
27	8	40.0	448	14	CD201120	CD201120 MS1-0124U
C 28	8	40.0	457	9	AV981198	AV981198 AV981198
C 29	8	40.0	458	28	BZ615844	BZ615844 1955D08.b
30	8	40.0	468	28	AQ945119	AQ945119 Sheared D
31	8	40.0	470	28	AQ702871	AQ702871 HS 5452 B
C 32	8	40.0	481	13	BM196597	BM196597 BM196597
33	8	40.0	490	9	AV890393	AV890393 AV890393
34	8	40.0	493	9	AV862905	AV862905 AV862905
C 35	8	40.0	501	13	BW202753	BW202753 BW202753
C 36	8	40.0	504	14	CD112744	CD112744 ME1-0024T
C 37	8	40.0	505	14	CA820687	CA820687 EST0012 N
C 38	8	40.0	510	9	AV950469	AV950469 AV950469
C 39	8	40.0	510	29	TA90C02P	AL465308 T. brucei
C 40	8	40.0	515	29	CC476294	CC476294 CH240_302
C 41	8	40.0	516	12	BP001077	BP001077 BP001077
C 42	8	40.0	521	13	BW136624	BW136624 BW136624
C 43	8	40.0	524	13	BW203262	BW203262 BW203262
C 44	8	40.0	525	9	AV680124	AV680124 AV680124
C 45	8	40.0	526	9	AV973566	AV973566 AV973566

ALIGNMENTS

RESULT 1
BF910533/c
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS

BF910533
CM4-UT0042-011100-395-f08 UT0042 Homo sapiens cDNA, mRNA sequence.
BF910533
BF910533.1 GI:12301991
Homo sapiens (human)
Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 240)
Dias Neto,E., Garcia Correa,R., Verjovskij-Almeida,S., Briones,M.R.,

Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F., Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H., Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and Simpson,A.J.

Shotgun sequencing of the human transcriptome with ORF expressed

sequence tags

Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)

20202663

10737800

COMMENT

Contact: Simpson A.J.G.

Laboratory of Cancer Genetics

Ludwig Institute for Cancer Research

Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,

Brazil

Tel: +55-11-2704922

Fax: +55-11-2707001

Email: asimpson@ludwig.org.br

This sequence was derived from the FAPESP/LICR Human Cancer Genome

Project. This entry can be seen in the following URL

(<http://www.ludwig.org.br/scripts/gethtml2.pl?tl=CM4&t2=CM4-UT0042-011100-395-f08&t3=2000-11-01&t4=1>)

Seq primer: puc 18 forward

High quality sequence start: 81

High quality sequence stop: 240.

Location/Qualifiers

FEATURES

source

1..240

/organism="Homo sapiens"

/mol_type="mRNA"

/db_xref="taxon:9606"

/dev_stage="Adult"

/clone_lib="UT0042"

/note="Organ: uterus_tumor; Vector: puc18; Site_1: Smar;

Site 2: Smal; A mini-library was made by cloning products

derived from ORESTES PCR (U.S. Letters Patent application

No. 196,716 - Ludwig Institute for Cancer Research)

profiles into the puc 18 vector. Reverse transcription of

tissue mRNA and cDNA amplification were performed under

low stringency conditions."

ORIGIN

Alignment Scores:
Pred. No.: 6,33e-09 Length: 240
Score: 19.00 Matches: 19
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 95.00% Indels: 0
DB: 10 Gaps: 0

US-10-799-747-116 (1-20) x BF910533 (1-240)

Qy 1 MetAlaAlaHisSerValLeuSerPheLeuLeuThrProTyAlaLeuLysSer 19

Db 191 ATGGCAGCCCAATTCAGTCTTGAGTTTCTCTCTGGACACCTTATGCTCTGAAATCA 135

RESULT 2

BM722991

LOCUS

DEFINITION

UI-E-EO1-aid-a-14-0-UI-r1 UI-E-EO1 Homo sapiens cDNA clone

UI-E-EO1-aid-a-14-0-UI 5', mRNA sequence.

BM722991

EST.

KEYWORDS

Source

ORGANISM

Homo sapiens (human)

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 575)

Bonaldi,M.F., Lennon,G. and Soares,M.B.

Normalization and subtraction: two approaches to facilitate gene

discovery

Genome Res. 6 (9), 791-806 (1996)

97044477

PUBMED

COMMENT

8889548

Contact: Soares, MB

Coordinated Laboratory for Computational Genomics

University of Iowa

375 Newton Road, 4156 MEBRF, Iowa City, IA 52242, USA

Tel: 319 335 8250

Fax: 319 335 9565

Email: bento-soares@uiowa.edu

Tissue Procurement: Dr. Gregg Hageman

cDNA Library preparation: Dr. M. Bento Soares, University of Iowa

cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa

DNA Sequencing by: Dr. M. Bento Soares, University of Iowa

Clone Distribution: Researchers may obtain clones from Research

Genetics (www.resgen.com)

The following repetitive elements were found in this cDNA

sequence: 148-169, >AT_RichLow_complexity (matched complement)

Seq primer: M13 Reverse.

Location/Qualifiers

source

1..575

/organism="Homo sapiens"

/mol_type="mRNA"

/db_xref="taxon:9606"

/clone="UI-E-EO1-aid-a-14-0-UI"

/tissue type="fetal eye"

/dev stage="fetal"

/lab_host="DH10B (Life Technologies) (T1 phage resistant)"

/clone_lib="UI-E-EO1"

/note="Organ: eye; Vector: pT7T3-Pac (Pharmacia) with a

modified polylinker; Site_1: EcoR I; Site 2: Not I;

UI-E-EO1 is a normalized cDNA library containing the

following tissue(s): fetal eye. The library was

constructed according to Bonaldi, Lennon and Soares,

Genome Research, 6:791-806, 1996. First strand cDNA

synthesis was primed with an oligo-dT primer containing a

Not I site. Double stranded cDNA was ligated to an EcoR I

adaptor, digested with Not I, and cloned directionally

into pT7T3-Pac vector. The oligonucleotide used to prime

the synthesis of first-strand cDNA contains a library tag

sequence that is located between the Not I site and the

(dT)18 tail. The sequence tag for this library is

CGCGTATACC. This library was created for the program, Gene

Discovery in the Visual System, supported by National Eye

Institute (NEI)."

ORIGIN

Alignment Scores:
Pred. No.: 1.55e-08 Length: 575
Score: 19.00 Matches: 19
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 95.00% Indels: 0
DB: 12 Gaps: 0

US-10-799-747-116 (1-20) x BM722991 (1-575)

Qy 1 MetAlaAlaHisSerValLeuSerPheLeuLeuThrProTyAlaLeuLysSer 19

Db 454 ATGGCAGCCCAATTCAGTCTTGAGTTTCTCTCTGGACACCTTATGCTCTGAAATCA 510

RESULT 3

BI914473

LOCUS

DEFINITION

BI914473

mRNA sequence.

BI914473

VERSION

BI914473.1 GI:16178652

KEYWORDS

EST.

SOURCE

Homo sapiens (human)

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 765)

NIH-MGC <http://mgc.nci.nih.gov/>.

BI914473

603182264F1 NIH_MGC_121 Homo sapiens cDNA clone IMAGE:5246463 5',

mRNA sequence.

BI914473

765 bp linear EST 16-OCT-2001

TITLE
JOURNAL
COMMENT

National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Life Technologies, Inc.
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM11621 row: p column: 16
High quality sequence stop: 762.

FEATURES
source
1. .765
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:5246463"
/lab_host="DH10B"
/clone_lib="NIH_MGC_121"
/note="Organ: brain; Vector: pCMV-SPORT6; Site 1: NotI; Site 2: EcoRV (destroyed); RNA source anonymous pool of 3 fetal brains, female age 20 weeks, female age 24 weeks, and male age 26 weeks. Library is oligo-dT primed and directionally cloned (EcoRV site is destroyed upon cloning). Average insert size 1.7 kb, insert size range 0.7-3.5 kb. Library is normalized and enriched for full-length clones and was constructed by C. Gruber (Invitrogen). Research Genetics tracking code 017. Note: this is a NIH_MGC Library."

ORIGIN

Alignment Scores:
Pred. No.: 2.07e-08 Length: 765
Score: 19.00 Matches: 19
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 95.00% Indels: 0
DB: 12 Gaps: 0

US-10-799-747-116 (1-20) x BI914473 (1-765)

Qy 1 MetAlaLaHisSerValLeuSerPheLeuLeuThrProTyrAlaLeuLysSer 19
|||||
Db 556 ATGCAGCCCATTCAGTCTGAGTTTCTCTCTGGACACCTATGCTCTGAATCA 612
|||||

RESULT 4
BG484396 785 bp mRNA linear EST 21-MAR-2001
LOCUS
DEFINITION 602505037F1 NIH_MGC_77 Homo sapiens cDNA clone IMAGE:4618473 5', mRNA sequence.
ACCESSION BG484396
VERSION BG484396.1 GI:13416675
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 785)
NIH-MGC http://mgc.nci.nih.gov/
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: CLONTECH Laboratories, Inc.
cDNA Library Preparation: CLONTECH Laboratories, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLCM1376 row: j column: 10

High quality sequence start: 18
High quality sequence stop: 751.

FEATURES
source
1. .785
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:4618473"
/lab_host="DH10B (T1 phage-resistant)"
/clone_lib="NIH_MGC_77"
/note="Organ: lung; Vector: pDNR-LIB (Clontech); Site 1: SfiI (ggccgctcgcc); Site 2: SfiI (ggccattggcc); 5' and 3' adaptors were used in cloning as follows: 5' adaptor sequence: 5'-CACGCCATTATGGC-3' and 3' adaptor sequence: 5'-ATTCTAGAGCGCGCGCGCATG-DT(30)BN-3' (where B = A, C, or G and N = A, C, G, or T). Average insert size 1.9 kb (range 0.5-4.0 kb). 12/15 colonies contained inserts by PCR. This library was enriched for full-length clones and was constructed by Clontech Laboratories (Palo Alto, CA). Note: this is a NIH_MGC Library."

ORIGIN

Alignment Scores:
Pred. No.: 2.13e-08 Length: 785
Score: 19.00 Matches: 19
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 95.00% Indels: 0
DB: 12 Gaps: 0

US-10-799-747-116 (1-20) x BG484396 (1-785)

Qy 1 MetAlaLaHisSerValLeuSerPheLeuLeuThrProTyrAlaLeuLysSer 19
|||||
Db 322 ATGCAGCCCATTCAGTCTGAGTTTCTCTCTGGACACCTATGCTCTGAATCA 378
|||||

RESULT 5
BX441923 1201 bp mRNA linear EST 15-MAY-2003
LOCUS
DEFINITION BX441923 Homo sapiens FETAL BRAIN Homo sapiens cDNA clone CSODF023Y002 5-PRIME, mRNA sequence.
ACCESSION BX441923
VERSION BX441923.1 GI:30771989
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 1201)
Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
Full-length cDNA libraries and normalization
Unpublished (2001)
Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 EVRY cedex - France
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
Library was constructed by Life Technologies, a division of
Invitrogen. This sequence belongs to sequence cluster 2532.r For
more information about this cluster, see
http://www.genoscope.cns.fr/
cgi-bin/cluster.cgi?seq=CSODF023BH01QPI&cluster=2532.r. Contact :
Feng Liang Email : fliang@lifetech.com URL :
http://fulllength.invitrogen.com/ Invitrogen Corporation 1600
Faraday Avenue Genoscope sequence ID : CSODF023BH01QPI.

FEATURES
source
1. .1201
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CSODF023Y002"
/tissue_type="FETAL BRAIN"
/dev_stage="Fetal"
/clone_lib="Homo sapiens FETAL BRAIN"

/note="Organ: brain; Vector: pCMVSPORT 6; 1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-strand cDNA was digested with Not I and cloned into the Not I and EcoRV sites of the pCMVSPORT 6 vector. Library was not normalized."

ORIGIN

Alignment Scores:
Pred. No.: 3.28e-08 Length: 1201
Score: 19.00 Matches: 19
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 95.00% Indels: 0
DB: 13 Gaps: 0

US-10-799-747-116 (1-20) x BX441923 (1-1201)

QY 1 MetAlaAlaHisSerValLeuSerPheLeuLeuTTPThrProTyAlaLeuLySer 19
Db ATGGCAGCCCATTCAGTCTTGAGTTTCTCTCTGGACACCTTATGCTCTGAAATCA 202

RESULT 6
AL035941
LOCUS DKZP564B1622 r1.564 (synonym: hfr2) Homo sapiens cDNA clone
DEFINITION DKZP564B1622 5', mRNA sequence.
ACCESSION AL035941
VERSION AL035941.1 GI:5405575
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
AUTHORS Wambutt, R., Heubner, D., Mewes, H.W., Gassenhuber, J. and Wiemann, S.
TITLES EST (Wambutt, et al.)
JOURNAL Unpublished (1999)
COMMENT Contact: MIPS

MIPS
Ingolstaedter Landstr.1, D-85764 Neuherberg, Germany
This is the 5' sequence of the clone insert
Clone from S. Wiemann, Molecular Genome Analysis, German Cancer Research Center (DKFZ); Email: s.wiemann@dkfz-heidelberg.de;
sequenced by AGOWA (Berlin/Germany) within the cDNA sequencing consortium of the German Genome Project.
SI sequence also available.

This clone (DKFZP564B1622) is available at the RZPD in Berlin.
Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059 Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de.

FEATURES

source
1..621
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="DKFZP564B1622"
/tissue_type="brain"
/dev_stage="fetal"
/lab_host="Xl-2blue"
/clone_lib="564 (synonym: hfr2)"
/note="Vector: pAMP1; Site_1: NotI; Site_2: SalI"

ORIGIN

Alignment Scores:
Pred. No.: 6.57 Length: 621
Score: 10.00 Matches: 10
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 50.00% Indels: 0
DB: 9 Gaps: 0

US-10-799-747-116 (1-20) x AL035941 (1-621)

QY 1 MetAlaAlaHisSerValLeuSerPheLeu 10

Db 591 ATGGCAGCCCATTCAGTCTTGAGTTTCTT 620

RESULT 7

AL0293004
LOCUS

DEFINITION HS_2225_A1_E02_MF CIT Approved Human Genomic Sperm Library D Homo sapiens genomic clone Plate=2225 Col=3 Row=I, genomic survey sequence.
ACCESSION AQ293004
VERSION AQ293004.1 GI:4010727
KEYWORDS GSS.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
AUTHORS 1 (bases 1 to 496)
Mahairas, G.G., Wallace, J.C., Smith, K., Swartzell, S., Holzman, T., Keller, A., Shaker, R., Furlong, J., Young, J., Zhao, S., Adams, M.D. and Hood, L.

TITLE Sequence-tagged connectors: A sequence approach to mapping and scanning the human genome

JOURNAL Proc. Natl. Acad. Sci. U.S.A. 96 (17), 9739-9744 (1999)
MEDLINE 99380589
PubMed 10449764
COMMENT Contact: Mahairas GG, Wallace JC, Hood L

High Throughput Sequencing Center
University of Washington
401 Queen Anne Avenue North, Seattle, WA 98109, USA
Tel: (206) 616-3618
Fax: (206) 616-3887
Email: jwallace@u.washington.edu

Sequence Tagged Connector
Place: 2225 row: I column: 3
Class: BAC ends
High quality sequence stop: 496.
Location/Qualifiers
1..496
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/clone="Plate=2225 Col=3 Row=I"
/sex="male"
/clone_lib="CIT Approved Human Genomic Sperm Library D"
/note="Organ: sperm; Vector: pBelobAC11; BAC Clones in E-Coli DH10B"

FEATURES

source
1..496
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/clone="Plate=2225 Col=3 Row=I"
/sex="male"
/clone_lib="CIT Approved Human Genomic Sperm Library D"
/note="Organ: sperm; Vector: pBelobAC11; BAC Clones in E-Coli DH10B"

ORIGIN

Alignment Scores:
Pred. No.: 47.1 Length: 496
Score: 9.00 Matches: 9
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 45.00% Indels: 0
DB: 28 Gaps: 0

US-10-799-747-116 (1-20) x AQ293004 (1-496)

QY 5 SerValLeuSerPheLeuLeuTTPThr 13

Db 119 TCTGTATTATCTCTCTGCTATGACT 145

RESULT 8

AV851019
LOCUS

DEFINITION AV851019 Nori Satoh unpublished cDNA library, larva Ciona intestinalis cDNA clone rcilvlon02 3', mRNA sequence.
ACCESSION AV851019
VERSION AV851019.1 GI:16834788
KEYWORDS EST.
SOURCE Ciona intestinalis
ORGANISM Ciona intestinalis

Eukaryota; Metazoa; Chordata; Urochordata; Ascidiacea; Enterogona;

AV851019 239 bp mRNA linear EST 08-NOV-2001
AV851019 Nori Satoh unpublished cDNA library, larva Ciona intestinalis cDNA clone rcilvlon02 3', mRNA sequence.

AV851019
ACCESSION AV851019
VERSION AV851019.1 GI:16834788
KEYWORDS EST.
SOURCE Ciona intestinalis
ORGANISM Ciona intestinalis

REFERENCE
1 (bases 1 to 239)
Satcho, N., Satcho, Y., Kohara, Y. and Shin-i, T.
Expressed genes in Ciona intestinalis
Unpublished (2000)
JOURNAL
COMMENT
Contact: Nori Satcho
Department of Zoology
Kyoto University
Sakyo-ku, Kyoto, Kyoto 606-8502, Japan
Tel: 81-75-753-4081
Fax: 81-75-705-1113
Email: satcho@ascidian.zool.kyoto-u.ac.jp.

FEATURES
source
Location/Qualifiers

1..239
/organism="Ciona intestinalis"
/mol_type="mRNA"
/db_xref="taxon:7719"
/clone="rcilv10n02"
/tissue_type="whole animal"
/dev_stage="larva"
/clone_lib="Nori Satcho unpublished cDNA library, larva"

ORIGIN
Alignment Scores:

Pred. No.: 201 Length: 239
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservatve: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 40.00% Indels: 0
DB: 9 Gaps: 0

US-10-799-747-116 (1-20) x AV851019 (1-239)

Qy 4 HisServValLeuSerPheLeu 11
|||||
Db 173 CATTGAGTTTGTCTTTTGTCTGTC 196

RESULT 9
BG371993
LOCUS
DEFINITION
UI-R-CVO-brl-a-01-0-UI.s1 UI-R-CVO Rattus norvegicus cDNA clone
UI-R-CVO-brl-a-01-0-UI 3', mRNA sequence.

ACCESSION
BG371993
VERSION
BG371993.1 GI:13268530

KEYWORDS
EST.

SOURCE
Rattus norvegicus (Norway rat)

ORGANISM
Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.

REFERENCE
1 (bases 1 to 272)

AUTHORS
Bonaldo, M.F., Lennon, G. and Soares, M.B.

TITLE
Normalization and subtraction: two approaches to facilitate gene

discovery

JOURNAL
Genome Res. 6 (9), 791-806 (1996)

MEDLINE
97044477

PUBMED
8889548

COMMENT
Contact: Soares, MB
Coordinated Laboratory for Computational Genomics
University of Iowa
375 Newton Road, 4156 MEBRF, Iowa City, IA 52242, USA
Tel: 319 335 8250
Fax: 319 335 9565
Email: bento-soares@uiowa.edu

The sequence contained an oligo-dT track that was present in the
oligonucleotide that was used to prime the synthesis of first
strand cDNA and therefore this may represent a bonafide poly A
tail. The sequence tag present in the cDNA between the NotI site
and the oligo-dT track served to verify it as a clone from the
non-normalized rat eye library cDNA Library Preparation: M.B.
Soares Lab Clone Distribution: clones will be available through
Research Genetics (www.resgen.com) The following repetitive
elements were found in this cDNA sequence: 1-25,

>AT rich#Low complexity
Seq primer: M13 Forward
POLYA=Yes.

FEATURES
source
Location/Qualifiers

1..272
/organism="Rattus norvegicus"
/mol_type="mRNA"
/strain="Sprague-Dawley"
/db_xref="taxon:10116"
/clone="UI-R-CVO-brl-a-01-0-UI"
/dev_stage="ADULT"
/lab_host="DH10B (Life Technologies)"
/clone_lib="UI-R-CVO"
/note="Vector: pT73D-Pac (Pharmacia) with a modified
polylinker; Site 1: Not I; Site 2: Eco RI; The UI-R-CVO
library is a non-normalized library constructed from rat
eye tissue. For a detailed description of the library from
which this clone was derived, please visit our web site at
rategest.eng.uiowa.edu. The subtraction has been previously
described in (Bonaldo, Lennon and Soares, Genome Research
6:791-806, 1996)
TAG_TISSUE=rat eye
TAG_LIB=UI-R-CVO
TAG_SEQ=CAGCC"

ORIGIN

Alignment Scores:
Pred. No.: 230 Length: 272
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservatve: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 40.00% Indels: 0
DB: 12 Gaps: 0

US-10-799-747-116 (1-20) x BG371993 (1-272)

Qy 2 AlaAlaHisServValLeuSerPhe 9
|||||

Db 107 GCAGCACATTCGTGCTTCTTC 130

RESULT 10

CE488576/c

LOCUS

DEFINITION

tigr-gss-dog-17000365567544 Dog Library Canis familiaris genomic,

genomic survey sequence.

ACCESSION

CE488576

VERSION

CE488576.1 GI:36805357

KEYWORDS

GSS.

SOURCE

Canis familiaris (dog)

ORGANISM

Canis familiaris

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.

REFERENCE

1 (bases 1 to 290)

AUTHORS

Kirkness, E.F., Bafna, V., Halpern, A.L., Levy, S., Remington, K.,

Rusch, D.B., Delcher, A.L., Pop, M., Wang, W., Fraser, C.M. and

Venter, J.C.

TITLE

The dog genome: survey sequencing and comparative analysis

JOURNAL

Science 301 (5641), 1898-1903 (2003)

MEDLINE

22875432

PUBMED

14512627

COMMENT

Contact: Kirkness EF

The Institute for Genomic Research

Department of Eukaryotic Genomics, TIGR, 9712 Medical Center Drive,

Rockville, MD 20850, USA

Tel: 301-838-0200

Fax: 301-838-0208

Email: ekirknes@tigr.org

Class: shotgun.

FEATURES

source

Location/Qualifiers

1..290

/organism="Canis familiaris"

/mol_type="genomic DNA"

/strain="Standard Poodle"

```

/db_xref="taxon:9615"
/clone_lib="Dog Library"
/note="Site 1: BstXI; Libraries were prepared from
peripheral blood"

ORIGIN
Alignment Scores:
Pred. No.: 245 Length: 290
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 40.00% Indels: 0
DB: 29 Gaps: 0

US-10-799-747-116 (1-20) x CE489576 (1-290)

Qy 4 HisSerValLeuSerPheLeu 11
Db 43 CACTCTGTCCTATCGTTCTACTG 20

RESULT 11
AQ350287/c
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
Homo sapiens (human)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
AUTHORS
Zhao,S., Adams,M.D., Nierman,W., Malek,J., de Jong,P. and
Venter,J.C.
TITLE
Use of BAC End Sequences from Library RPCI-11 for Sequence-Ready
Map Building
JOURNAL
Unpublished (1997)
COMMENT
Contact: Shaying Zhao, William Nierman, Mark Adams
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850
Tel: 301 838 0200
Fax: 301 838 0208
Email: hbs@igr.org
Clones are derived from the human BAC library RPCI-11. For BAC
library availability, please contact Pieter de Jong
(pieter@dejong.med.buffalo.edu). Clones may be purchased from
BACPAC Resources (http://bacpac.med.buffalo.edu/ordering) or from
Research Genetics (info@resgen.com). BAC end search page:
http://www.tigr.org/tdb/hungen/bac\_end\_search/bac\_end\_search.html
Seq primer: SP6
Class: BAC ends.
FEATURES
Location/Qualifiers
source
1..323
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="GDB:7551006"
/db_xref="taxon:9606"
/clone="RPCI-11-133N7"
/sex="Male"
/cell_type="Lymphocytes"
/clone_lib="RPCI-11"
/note="Vector: pBACe3.6; Site_1: EcoRI; Site_2: EcoRI;
RPCI11 Human Male BAC Library"

ORIGIN
Alignment Scores:
Pred. No.: 274 Length: 323
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 40.00% Indels: 0
DB: 29 Gaps: 0

US-10-799-747-116 (1-20) x AQ350287 (1-323)

Qy 4 HisSerValLeuSerPheLeu 11
Db 137 CATTCCGTCCTCTCTCTCTCTCTT 114

RESULT 12
CC808823/c
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
Zea mays subsp. mays (maize)
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
clade; Panicoideae; Andropogoneae; Zea.
REFERENCE
AUTHORS
Bharti,A.K., Young,S., Kavchok,S., Keizer,G., Bronzino,A.C.,
Rouzaud,K., Fuks,G., Yu,Y., Wang,R. and Messing,J.
TITLE
Sequencing of the maize genome at PGIR (2003b)
JOURNAL
Unpublished (2003)
COMMENT
Contact: Bharti,A.K.
Dr.Joachim Messing's lab
The Plant Genome Initiative at Rutgers, Waksman Institute, Rutgers
University
190 Frelinghuysen Road, Piscataway, NJ 08854, USA
Tel: 732 445 3801
Fax: 732 445 5735
Email: bharti@waksman.rutgers.edu
Seq primer: SP6
Class: BAC ends
High quality sequence start: 88.
FEATURES
Location/Qualifiers
source
1..328
/organism="Zea mays subsp. mays"
/mol_type="genomic DNA"
/cultivar="B73"
/sub_species="mays"
/db_xref="taxon:4578"
/clone="ZMMBBc0476P22"
/lab_host="E. coli DH10B"
/clone_lib="ZMMBBc"
/note="Vector: p7ARBAC1.3; Site_1: BamHI; Site_2: BamHI"

ORIGIN
Alignment Scores:
Pred. No.: 278 Length: 328
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 40.00% Indels: 0
DB: 29 Gaps: 0

US-10-799-747-116 (1-20) x CC808823 (1-328)

Qy 3 AlaHisSerValLeuSerPheLeu 10
Db 35 GCTCACTCTGCTCTTAGTTTGTG 12

RESULT 13
CD202965
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
Schistosoma mansoni
CD202965
MS1-0139P-V386-G01-U.B MSI-0139 Schistosoma mansoni cDNA clone
MS1-0139P-V386-G01.B, mRNA sequence.
CD202965
EST:
Schistosoma mansoni
CD202965
336 bp mRNA linear EST 14-S3P-2003

```

```

DB: 28 Gaps: 0

US-10-799-747-116 (1-20) x AQ350287 (1-323)

Qy 4 HisSerValLeuSerPheLeu 11
Db 137 CATTCCGTCCTCTCTCTCTCTT 114

RESULT 12
CC808823/c
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
Zea mays subsp. mays (maize)
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
clade; Panicoideae; Andropogoneae; Zea.
REFERENCE
AUTHORS
Bharti,A.K., Young,S., Kavchok,S., Keizer,G., Bronzino,A.C.,
Rouzaud,K., Fuks,G., Yu,Y., Wang,R. and Messing,J.
TITLE
Sequencing of the maize genome at PGIR (2003b)
JOURNAL
Unpublished (2003)
COMMENT
Contact: Bharti,A.K.
Dr.Joachim Messing's lab
The Plant Genome Initiative at Rutgers, Waksman Institute, Rutgers
University
190 Frelinghuysen Road, Piscataway, NJ 08854, USA
Tel: 732 445 3801
Fax: 732 445 5735
Email: bharti@waksman.rutgers.edu
Seq primer: SP6
Class: BAC ends
High quality sequence start: 88.
FEATURES
Location/Qualifiers
source
1..328
/organism="Zea mays subsp. mays"
/mol_type="genomic DNA"
/cultivar="B73"
/sub_species="mays"
/db_xref="taxon:4578"
/clone="ZMMBBc0476P22"
/lab_host="E. coli DH10B"
/clone_lib="ZMMBBc"
/note="Vector: p7ARBAC1.3; Site_1: BamHI; Site_2: BamHI"

ORIGIN
Alignment Scores:
Pred. No.: 278 Length: 328
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 40.00% Indels: 0
DB: 29 Gaps: 0

US-10-799-747-116 (1-20) x CC808823 (1-328)

Qy 3 AlaHisSerValLeuSerPheLeu 10
Db 35 GCTCACTCTGCTCTTAGTTTGTG 12

RESULT 13
CD202965
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
Schistosoma mansoni
CD202965
MS1-0139P-V386-G01-U.B MSI-0139 Schistosoma mansoni cDNA clone
MS1-0139P-V386-G01.B, mRNA sequence.
CD202965
EST:
Schistosoma mansoni
CD202965
336 bp mRNA linear EST 14-S3P-2003

```

ORGANISM Schistosoma mansoni
Eukaryota; Metazoa; Platyhelminthes; Trematoda; Digenea;
Strigeiida; Schistosomatoidea; Schistosomatidae; Schistosoma.
1 (bases 1 to 336)

REFERENCE
AUTHORS Verjovski-Almeida, S., DeMarco, R., Martins, E.A.L., Guimaraes, P.E.M.,
Ojopi, E.P.B., Paquola, A.C.M., Piazza, J.P., Nishiyama, M.Y. Jr.,
Kitajima, J.P., Adamson, R.E., Ashton, P.D., Bonaldo, M.F.,
Coulson, P.S., Dillon, G.P., Farias, L.P., Gregorio, S.P., Ho, P.L.,
Leite, R.A., Malaquias, L.C.C., Marques, R.C.P., Miyasato, P.A.,
Nascimento, A.L.T.O., Ohlweiler, F.P., Reis, E.M., Ribeiro, M.A.,
Sa, R.G., Stukart, G.C., Soares, M.B., Gargioni, C., Kawano, T.,
Rodrigues, V., Madeira, A.M.B.N., Wilson, R.A., Menck, C.F.M.,
Setubal, J.C., Leite, L.C.C. and Dias-Neto, E.

TITLE Transcriptome analysis of the acclimated human parasite Schistosoma
mansoni

JOURNAL Nat. Genet. 35 (2), 148-157 (2003)
MEDLINE 22879926
COMMENT Contact: Dr. Sergio Verjovski-Almeida
Departamento de Bioquímica
Instituto de Química - Universidade de São Paulo
Av. Prof. Lineu Prestes 748 sala 1200, 05508-900 São Paulo - SP,
Brasil
Tel: +55-11-3091-2173
Fax: +55-11-3091-2186
Email: verjo@iq.usp.br

This sequence was derived from the FAPESP Schistosoma mansoni EST
Genome Project. All sequences in the project were assembled and
annotated. This entry and all the assembled sequences can be seen
in the following URL <http://bioinfo.iq.usp.br/schisto/>
Plate: MSI-0139P-V386 row: 1 column: G.

FEATURES
Location/Qualifiers
source
1..336
/organism="Schistosoma mansoni"
/mol_type="mRNA"
/db_xref="taxon:6183"
/clone="MSI-0139P-V386-G01.B"
/sex="mixed pool"
/dev_stage="schistosomulum"
/lab_host="in vitro culture"
/clone_lib="MSI-0139"
/note="Vector: pGEM T-easy"

ORIGIN
Alignment Scores:
Pred. No.: 285 Length: 336
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 40.00% Indels: 0
DB: 14 Gaps: 0

US-10-799-747-116 (1-20) x CD202965 (1-336)

Qy 4 HisSerValLeuSerPheLeuLeu 11
|||||
Db 99 CATTGGGTTTAAAGTTTCTCTG 122
|||||

RESULT 14
T59368
LOCUS yb57h01.s1 Stragatene ovary (#937217) Homo sapiens cDNA clone
DEFINITION IMAGE:75313 3' similar to gb:X77738 rnal BAND 3 ANION TRANSPORT
PROTEIN (HUMAN) ; contains Alu repetitive element; , mRNA sequence.
T59368
ACCESSION T59368.1 GI:661205
VERSION T59368.1
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
AUTHORS Hillier, L., Lennon, G., Becker, M., Bonaldo, M.F., Chiapelli, B.,
Chissoe, S., Dietrich, N., DuBuque, T., Favellio, A., Gish, W.,

ORGANISM Schistosoma mansoni
Eukaryota; Metazoa; Platyhelminthes; Trematoda; Digenea;
Strigeiida; Schistosomatoidea; Schistosomatidae; Schistosoma.
1 (bases 1 to 336)

REFERENCE
AUTHORS Verjovski-Almeida, S., DeMarco, R., Martins, E.A.L., Guimaraes, P.E.M.,
Ojopi, E.P.B., Paquola, A.C.M., Piazza, J.P., Nishiyama, M.Y. Jr.,
Kitajima, J.P., Adamson, R.E., Ashton, P.D., Bonaldo, M.F.,
Coulson, P.S., Dillon, G.P., Farias, L.P., Gregorio, S.P., Ho, P.L.,
Leite, R.A., Malaquias, L.C.C., Marques, R.C.P., Miyasato, P.A.,
Nascimento, A.L.T.O., Ohlweiler, F.P., Reis, E.M., Ribeiro, M.A.,
Sa, R.G., Stukart, G.C., Soares, M.B., Gargioni, C., Kawano, T.,
Rodrigues, V., Madeira, A.M.B.N., Wilson, R.A., Menck, C.F.M.,
Setubal, J.C., Leite, L.C.C. and Dias-Neto, E.

TITLE Transcriptome analysis of the acclimated human parasite Schistosoma
mansoni

JOURNAL Nat. Genet. 35 (2), 148-157 (2003)
MEDLINE 22879926
COMMENT Contact: Dr. Sergio Verjovski-Almeida
Departamento de Bioquímica
Instituto de Química - Universidade de São Paulo
Av. Prof. Lineu Prestes 748 sala 1200, 05508-900 São Paulo - SP,
Brasil
Tel: +55-11-3091-2173
Fax: +55-11-3091-2186
Email: verjo@iq.usp.br

This sequence was derived from the FAPESP Schistosoma mansoni EST
Genome Project. All sequences in the project were assembled and
annotated. This entry and all the assembled sequences can be seen
in the following URL <http://bioinfo.iq.usp.br/schisto/>
Plate: MSI-0139P-V386 row: 1 column: G.

FEATURES
Location/Qualifiers
source
1..336
/organism="Schistosoma mansoni"
/mol_type="mRNA"
/db_xref="taxon:6183"
/clone="MSI-0139P-V386-G01.B"
/sex="mixed pool"
/dev_stage="schistosomulum"
/lab_host="in vitro culture"
/clone_lib="MSI-0139"
/note="Vector: pGEM T-easy"

ORIGIN
Alignment Scores:
Pred. No.: 285 Length: 336
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 40.00% Indels: 0
DB: 14 Gaps: 0

US-10-799-747-116 (1-20) x CD202965 (1-336)

Qy 4 HisSerValLeuSerPheLeuLeu 11
|||||
Db 99 CATTGGGTTTAAAGTTTCTCTG 122
|||||

RESULT 14
T59368
LOCUS yb57h01.s1 Stragatene ovary (#937217) Homo sapiens cDNA clone
DEFINITION IMAGE:75313 3' similar to gb:X77738 rnal BAND 3 ANION TRANSPORT
PROTEIN (HUMAN) ; contains Alu repetitive element; , mRNA sequence.
T59368
ACCESSION T59368.1 GI:661205
VERSION T59368.1
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
AUTHORS Hillier, L., Lennon, G., Becker, M., Bonaldo, M.F., Chiapelli, B.,
Chissoe, S., Dietrich, N., DuBuque, T., Favellio, A., Gish, W.,

Hawkins, M., Hultman, M., Kucaba, T., Lacy, M., Le, N.,
Mardis, E., Moore, B., Morris, M., Parsons, J., Prange, C., Rifkin, L.,
Rohlfing, T., Schellenberg, K., Soares, M.B., Tan, F., Thierry-Mieg, J.,
Treviski, E., Underwood, K., Wohlmann, P., Waterston, R., Wilson, R.
and Marra, M.
Generation and analysis of 280,000 human expressed sequence tags
Genome Res. 6 (9), 807-828 (1996)
97044478
8889549
PUBMED
COMMENT
Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@wustl.wustl.edu
High quality sequence stops: 339
Source: IMAGE Consortium, LLNL
This clone is available royalty-free through LLNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Seq primer: -21ml3
High quality sequence stop: 339.

FEATURES

Location/Qualifiers

source

1..356
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="GDB:505042"
/db_xref="taxon:9606"
/clone="IMAGE:75313"
/sex="female"
/dev_stage="49 year old"
/lab_host="SOLR cells (kanamycin resistant)"
/clone_lib="Stratagene ovary (#937217)"
/note="Organ: ovary; Vector: Bluescript SK; Site 1: EcoRI;
Site 2: XhoI; Cloned unidirectionally. Primer: Oligo dt.
Total ovary tissue, normal, caucasian. Average insert
size: 0.8 Kb; Uni-ZAP XR Vector; -5' adaptor sequence: 5'
GAATTCGGCAGCAG 3' -3' adaptor sequence: 5'
CTCGAGTTTTTTTTTTTTTTT 3'"

ORIGIN

Alignment Scores:

Pred. No.: 302 Length: 356
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 40.00% Indels: 0
DB: 14 Gaps: 0

US-10-799-747-116 (1-20) x T59368 (1-356)

Qy 3 AlahisSerValLeuSerPheLeu 10

|||||
Db 108 GTCATCTCTGACTTCTCTCTC 131
|||||

RESULT 15

CE286752

LOCUS

DEFINITION

CE286752

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

22875432

PUBMED 14512627
COMMENT Contact: Kirkness EF
The Institute for Genomic Research
Department of Eukaryotic Genomics, TIGR, 9712 Medical Center Drive,
Rockville, MD 20850, USA
Tel: 301-838-0200
Fax: 301-838-0208
Email: ekirknes@tigr.org
Class: shotgun.
FEATURES Location/Qualifiers
source 1..385
/organism="Canis familiaris"
/mol_type="genomic DNA"
/strain="Standard Poodle"
/db_xref="taxon:9615"
/clone_lib="Dog Library"
/note="Site 1: BstXI; Libraries were prepared from
peripheral blood"

ORIGIN

Alignment Scores:
Pred. No.: 327 Length: 385
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 40.00% Indels: 0
DB: 29 Gaps: 0

US-10-799-747-116 (1-20) x CE286752 (1-385)

QY 4 HisSerValLeuSerPheLeuLeu 11
DB 198 CACAGCGTCTATCAATTCCTCTG 221

Search completed: July 21, 2004, 05:22:32
Job time : 2475 secs